



## Effective grain size distribution analysis for interpretation of tidal-deltaic facies: West Bengal Sundarbans

Flood, R. P., Orford, J. D., McKinley, J. M., & Roberson, S. (2015). Effective grain size distribution analysis for interpretation of tidal-deltaic facies: West Bengal Sundarbans. *Sedimentary Geology*, 318, 58-74. DOI: 10.1016/j.sedgeo.2014.12.007

**Published in:**  
Sedimentary Geology

**Document Version:**  
Peer reviewed version

**Queen's University Belfast - Research Portal:**  
[Link to publication record in Queen's University Belfast Research Portal](#)

### **Publisher rights**

Copyright © 2015 Elsevier.

This is an open access article published under a Creative Commons Attribution-NonCommercial-NoDerivs License (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits distribution and reproduction for non-commercial purposes, provided the author and source are cited.

### **General rights**

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### **Take down policy**

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [openaccess@qub.ac.uk](mailto:openaccess@qub.ac.uk).

## Accepted Manuscript

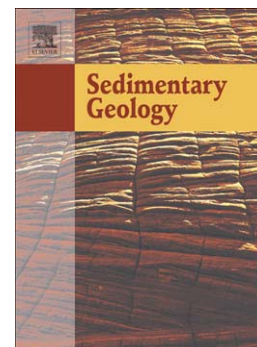
Effective grain size distribution analysis for interpretation of tidal-deltaic facies: West Bengal Sundarbans

R.P. Flood, J.D. Orford, J.M. McKinley, S. Roberson

PII: S0037-0738(15)00002-0  
DOI: doi: [10.1016/j.sedgeo.2014.12.007](https://doi.org/10.1016/j.sedgeo.2014.12.007)  
Reference: SEDGEO 4809

To appear in: *Sedimentary Geology*

Received date: 13 October 2014  
Revised date: 23 December 2014  
Accepted date: 26 December 2014



Please cite this article as: Flood, R.P., Orford, J.D., McKinley, J.M., Roberson, S., Effective grain size distribution analysis for interpretation of tidal-deltaic facies: West Bengal Sundarbans, *Sedimentary Geology* (2015), doi: [10.1016/j.sedgeo.2014.12.007](https://doi.org/10.1016/j.sedgeo.2014.12.007)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**Effective grain size distribution analysis for interpretation of tidal-deltaic facies:  
West Bengal Sundarbans**

R.P. Flood<sup>1\*</sup>, J.D. Orford<sup>1</sup>, J.M. McKinley<sup>1</sup>, S. Roberson<sup>2</sup>.

<sup>1</sup>School of Geography, Archaeology and Palaeoecology, Queen's University Belfast,  
Belfast, B7 1NN, Northern Ireland, UK

<sup>2</sup>Geological Survey of Northern Ireland, Colby House, Stranmillis Court, Belfast,  
BT9 5BF, Northern Ireland, UK

\*Corresponding author: rflood02@qub.ac.uk Phone number: +442890973929

**Keywords:** grain size distributions; compositional data; multivariate statistics;  
lithofacies; Sundarbans; Ganges-Brahmaputra delta.

**Abstract**

Research over the past two decades on the Holocene sediments from the tide dominated west side of the lower Ganges delta has focussed on constraining the sedimentary environment through grain size distributions (GSD). GSD has traditionally been assessed through the use of probability density function (PDF) methods (e.g. log-normal, log skew-Laplace functions), but these approaches do not acknowledge the compositional nature of the data, which may compromise outcomes in lithofacies interpretations. The use of PDF approaches in GSD analysis poses a

series of challenges for the development of lithofacies models, such as equifinal distribution coefficients and obscuring the empirical data variability. In this study a methodological framework for characterising GSD is presented through compositional data analysis (CODA) plus a multivariate statistical framework. This provides a statistically robust analysis of the fine tidal estuary sediments from the West Bengal Sundarbans, relative to alternative PDF approaches.

## 1. Introduction

This paper presents a methodological framework for characterising grain size variability in sediments through compositional data analysis (CODA) associated with a multivariate statistical framework (PCA and Cluster analysis). Grain size distributions (GSDs) consist of data that sum to a closed amount (i.e., 100%) and as a result only carry relative information with comparisons of such percentage data being considered to be spurious (Pearson, 1896; Chayes, 1960, 1971; Aitchison, 1986). This problem of closed data has been first highlighted by Pearson (1896) and Chayes (1960; 1971) and has led to the development of a series of log-ratio transformation by Aitchison (1986). The use of probability density functions (PDFs) for grain size analysis also poses a series of challenges, particularly where GSDs are not normally distributed (e.g., Friedman, 1962; Bagnold and Barndorff-Nielsen, 1980; Fredlund et al., 2000; Fieller et al., 1992; Beierle et al., 2002). Furthermore, challenges to data analysis through these PDF approaches may be present where there is multimodality in the distribution fitted which may lead to spurious determination of mean and variance using the a model distribution coefficients (Roberson and Weltje, 2014). Along with this, the use of PDFs to examine GSDs may actually obscure the underlying variability in a GSD dataset (Roberson and

Weltje, 2014). The primary aim of this paper is to provide statistically robust procedures for grain size analysis of fine tidal estuary sediments using an example of the West Bengal Sundarbans, through multivariate statistics (PCA and cluster analyses) along with CODA approaches. Two different methods to consider sediments from this same area have been presented: (i) by facies analysis as advocated by Allison et al. (2003); and (ii) the use of a traditional approach of grain size analysis (a log-normal product method of moments (Krumbein and Pettijohn, 1938; Friedman and Johnson, 1982) and a log skew-Laplace distribution (Fieller et al., 1984; Fieller and Flenley, 1992)) as advocated by Purkait and Majumdar (2014).

The methods applied in this paper involve decomposing grain size variability into ‘signals’ while firstly, overcoming the criteria imposed by the unit-sum constraint of closed system data (Chayes, 1960, 1971) in establishing the compositional nature of grain size data; and secondly the limitations of non-uniqueness associated with probability density functions (PDF) such as the log-normal and log skew-Laplace approaches when used with GSD. The approaches proposed in this paper have been applied for GSD interpretation of a Holocene sedimentary core (c. 6.5 m; 4,000 yr BP) from Lothian Island (a tidal island in the western Ganges Delta; Figs. 1, 2). The data generated from each of the different methods (i.e., log-ratio transformed data, log-normal and log skew-Laplace coefficients) will be assessed through multivariate statistical analyses in order to derive data-model comparisons along with a model intercomparison.

## **2. Analysis of grain size distributions in the Sundarbans**

### **2.1 Rationale for previous grain size approaches**

Allison et al. (2003) have characterised the stratigraphical facies of the lower Ganges-Brahmaputra delta plain as consisting of a fining upward sequence from a Muddy Sand facies (MS) to a Mottled Mud facies (MM) (or Interbedded Mud Facies (IMF)), with a final Peaty Mud facies (PM) located primarily in inland areas. The Holocene age sediments of the lower delta plain tend to be homogenous silts to clayey silts, with the Sundarbans to the west of the delta plain presenting slightly finer material on average in comparison to that of the eastern delta (Goodbred and Kuehl, 2000; Allison et al., 2003). Sand composition is less than 6% in these facies, but tends to decrease upcore as the MS facies grade into the fine silts and clayey silts (Allison et al., 2003). The modal class of sediments range between 6.6-7.6  $\phi$  for medium silts with the presence of a 3.6-4  $\phi$  modal class in samples possessing significant sand content (Allison et al., 2003). Although this analysis of grain size on the lower Ganges Brahmaputra (G-B) delta plain provides some context as to the overall facies sequence present; it offers no quantitative substance on the composition of facies sequences.

Purkait and Majumdar (2014) have attempted to analyse GSD from surface sediments of the Sundarbans through the use of log-normal and log skew-Laplace distribution analysis, and then discriminant function analysis of pertinent characterising indices of sediment grain size distribution analysis. The basis of applying these approaches, particularly the log-log approaches, is that particular percentile size and mean size discrimination are regarded as poor indicators of depositional process (Passega, 1957, 1964), while the use of dimensionality reduction approaches such as discriminant analysis (e.g., Sahu, 1964; Moiola and Spencer, 1979; Kasper-Zubillaga and Carranza-Edwards, 2005) and factor analysis (e.g., Klován, 1966) are also flawed in their relative efficacy to characterise such

environments (Purkait and Majumdar, 2014). Purkait and Majumdar (2014) have used graphical moment measurements in the log-normal distribution to derive the textural parameters of mean, sorting, skewness, and kurtosis after Krumbein (1934) and Folk and Ward (1957). Patterns in the grain size distribution reflecting different depositional environments were then compared using two model distributions: the log-normal and log skew-Laplace, in terms of discriminating inter- and sub-environmental groups of depositional environments (Purkait and Majumdar, 2014). The performance of discriminant function analysis using mean size, sorting, skewness, kurtosis, alpha, and beta from log-normal and log skew-Laplace are considered to be fit for purpose in discriminating textural facies (Purkait and Majumdar, 2014). The primary aim of this paper is to supplement these sedimentary-facies interpretations through a series of more statistically robust approaches to GSD from the Ganges delta-plain in general, and the Sundarbans in particular.

## 2.2 Grain size distributions: data constraints and solutions

The problems with these probability density function approaches advocated by Purkait and Majumdar (2014) have been outlined by Weltje and Prins (2007). Firstly, grain size distributions tend not to actually follow a log-normal distribution as has been reported by several authors (Friedman, 1962; Bagnold and Barndorff-Nielsen, 1980; Fredlund et al., 2000; Fieller et al., 1992; Beierle et al., 2002). An illustration (Roberson and Weltje, 2014) of this non-log-normal distribution in GSD is shown in Fig. 3, where despite multimodality in the sediments, their GSD can possess the same values of mean and variance using the log-normal distribution coefficients. This problem of equifinal distribution coefficients also applies to alternative distribution functions such as the log-hyperbolic and the log skew-Laplace

(Roberson and Weltje, 2014). Secondly, the application of PDFs to characterise grain size distributions have been found to conceal the actual variability present in the data (Roberson and Weltje, 2014).

An approach that may be applied to differentiating GSD variability, while avoiding these challenges posed by probability density functions, is to apply some form of multivariate statistical technique, most prominently eigenvector analysis in the form of principal components analysis (PCA) (Syvitski, 1991; Stauble and Cialone, 1996; Roberson and Weltje, 2014). The premise underlying the application of multivariate statistical techniques is that GSDs contain a proportion of mass in each size class bin (Weltje and Prins, 2003). Thus, these mass partitions relate to some attribute of a multivariate observation, as opposed to that of the GSD as belonging to some form of continuous function. In this approach, each GSD is considered to be a single datum comprising of as many components as there are size classes, with the multivariate approach designed to cover the relationships that exist between observations that comprise the dataset (Weltje and Prins, 2003). The key benefit of using multivariate approaches to GSD is that functional forms of groups, clusters or end-members are not specified prior to the analysis being performed (Weltje and Prins, 2003). Thus, when considering the shape of the GSD and the subpopulations contained within, one might consider multivariate approaches as being non-parametric in this regard (Weltje and Prins, 2003).

The key multivariate approaches developed for GSD analysis include that of cluster analysis (Zhou et al., 1991), entropy analysis (Forrest and Clark, 1989), factor analysis (Klovan, 1966; Solohub and Klovan, 1970; Allen et al., 1971, 1972; Dal Cin, 1976; Samthein et al., 1981; East, 1985, 1987; Syvitski, 1991) and finally principal components analysis, being one of the most popular approaches used for



these purposes (Davis, 1970; Chambers and Upchurch, 1979; Lirer and Vinci, 1991; Zhou et al., 1991). However, the effectiveness of these PDF approaches to assess grain size variability are somewhat compromised as the GSD data are compositional and constrained as data sum to 100%. Purkait and Majumdar (2014) offer a more numerically-based approach to Ganges delta-plain sediments in comparison to the approaches of Allison et al. (2003). However, as a result of the inherent problems surrounding the probability density function approaches and given the compositional nature of the data, a series of analytical approaches grounded in compositional data analysis (CODA) need to be undertaken.

## 2.3 Compositional Data Analysis

### 2.3.1 Unit-sum constraint and the limitations of constrained data

The primary drawback in the analysis of grain size data in any multivariate statistical scheme is that such data are compositional in nature and are thus constrained and cannot vary independently. Compositional data consist of any data set which sum to a constant sum as in a sediment distribution where relative weight in each size division is summed to 100% over the size range (i.e., data were measured in parts per unit, or 100% if measured in percentages), hence these data are sometimes referred to as 'closed' (Chayes, 1971). Compositional data can be seen as parts of a whole, in which only relative information is conveyed by the data (Pawlowsky-Glahn and Egozcue, 2006). In closed datasets, variables are not able to vary independently of one another and this can manifest itself in their variance-covariance structure (Aitchison, 1986). In the constant-sum constraint, there is at the very least, one forced covariance and one correlation coefficient between elements that will be negative (Pawlowsky-Glahn and Egozcue, 2006). With this one correlation being

negative, then the correlation coefficients between elements are thus not able to freely range between -1 and +1. As a result the data are constrained, with spurious correlations being induced as a result of the data summing to a constant and the data being closed with a bias present in the preference towards negative correlation (Pawlowsky-Glahn and Egozcue, 2006). In order to overcome the compositional data constraints, a framework of techniques founded on the principle of applying log-ratio transformation to data was developed by Aitchison (1986).

### 2.3.2 Compositional data analysis framework

There are two main transformations developed by Aitchison (1982): the additive-log-ratio (alr) and centred-log-ratio (clr) transformations. A composition therefore can now be represented as a real vector using these transformations (Pawlowsky-Glahn and Egozcue, 2006). The focus of this paper is on the application of the clr-transformation, which applies the geometric mean as the denominator,  $\text{clr}: S^D \rightarrow U^D$ :

$$z = \text{clr}(\mathbf{x}) = \left( \log \frac{x_1}{g(\mathbf{x})}, \log \frac{x_2}{g(\mathbf{x})}, \log \frac{x_D}{g(\mathbf{x})} \right) \quad (1)$$

Where:

$$U^D = \{[U_1, \dots, U_D] : U_1 + \dots + U_D = 0\} \quad (2)$$

is a hyperplane of  $\mathbb{R}^D$ , with the inverse transformation:  $\text{clr}^{-1}: U^D \rightarrow S^D$  in the form of:

$$\mathbf{x} = \text{clr}^{-1}(z) = c[\exp(z_1), \dots, \exp(z_D)] \quad (3)$$

Data from the clr-transformation is situated on a plane in D-dimensional real space, similar to that of data situated on a plane in compositional space (Pawlowsky-Glahn and Egozcue, 2006; Bloemsma, 2010). By contrast to compositional space, the

plane in real space is a hyper-plane, in which the solution space stretches infinitely in all directions (Bloemsma, 2010). This transformation allows for statistical analysis of compositional data to be carried out through multivariate approaches, as the data are now unconstrained data (Aitchison, 1982, 1986).

## **2.4 Multivariate statistical approaches to grain size analysis**

### **2.4.1 Principal Components Analysis**

Principal components analysis (PCA) is made up of a linear transformation of  $m$  original variables into  $m$ , new variables (components) where each new component is a linear combination of the original  $m$  variables (Davis, 1986). PCA is performed in a manner in which the total variance of the dataset is accounted for by each successive new component. PCA is not a true statistical approach to data analysis, but is a form of arithmetic dimensionality-reduction using singular value decomposition (SVD) of a data matrix (Van den Boogaart and Tolosana-Delgado, 2013). It should be noted here that the clr-transformation is consistent with the principle of least-squares estimation which also underlies the SVD approach. Furthermore, the number of non-zero eigenvectors of a clr-transformed data set equals  $D-1$  because of the sum-to-zero constraint on the clr vectors. The SVD approach deals with the computation of the principal components and eigenvalues of the initial data matrix through calculations based on either the covariance or correlation matrix. PCA involves the decomposition of the data into a matrix of ‘components’, from which, one may interpret the distributional shapes that make-up the overall GSD in the data (Weltje and Prins, 2003). A matrix of ‘loadings’ is computed that represent the extent to which each of the input GSD matches each of the components (Jöreskog et al., 1976; Davis, 1986; Weltje and Prins, 2003). From these loadings, principal component

scores are computed in which the original data are examined relative to the principal components calculated (Swan and Sandilands, 1995). In the following analysis, the interpretation of the properties of subpopulations is based on the eigenvectors calculated and the relative weighting of the scalar quantities of the eigenvalues.

#### 2.4.2 Cluster analysis

The key objective of cluster analysis (CA) is to partition a sequence of multivariate observations into more interpretable homogeneous groups (Sneath and Sokal, 1973; Templ et al., 2008). Using these derived groups provides a better understanding of the data structure and similarities between samples as well as differences between groups (Templ et al., 2008). One of the principal approaches in CA is that it is designed to minimise the variability of samples collected within a cluster group, whilst maximising the variability between groups (Templ et al., 2008). In order to determine what group membership exists, the majority of clustering methods employ some measure of similarity between observations, with Euclidean (straight line) distance between group centres being the numerical basis of similarity (Templ et al., 2008). The linkage approaches applied in this paper consist of the Ward's method (sample agglomeration) and the partitional k-means clustering method. Ward's method, also known as the incremental sum of squares method, uses both the squared distances of the within-cluster and between-cluster (Ward, 1963; Wishart 1969; Rencher, 2002). The use of the k-means algorithm is to derive a partition in cluster membership, in such a manner that the squared error, i.e., the error that exists between the empirical mean of a cluster and the points in the cluster, is minimised (Jain, 2010).

### 3 Grain size analysis of the Sundarbans: Data acquisition and treatments

#### 3.1 Sundarbans sediment retrieval

Shallow percussion cores of depths ranging between 3.5 and 7.0 m were obtained from a series of field sites in the Sundarbans (Flood, 2014). Data presented in this paper comes from the Lothian Island site (shown in Fig. 2). Sediment samples ( $n=83$ ) were collected from the open-split Lothian core at 8 cm intervals. Radiocarbon age estimation of the Lothian Island core is  $4291 \pm 35$   $^{14}\text{C}$  yr BP (c. 4231 – 4451 cal yr BP) at the core base, with the overall sedimentation rate estimated at c.  $1.4 \text{ mm a}^{-1}$  (Flood, 2014).

#### 3.2 Laser granulometry

Grain size distributions were obtained using the optical laser method of the MalvernMastersizer 2000 instrument (Malvern Instruments, Malvern, England). Following collection of measurements from the MalvernMastersizer 2000, data were aggregated into quarter phi intervals ( $\phi$  scale) over the range of  $0.02 - 2000 \mu\text{m}$ .

#### 3.3 Implementation of statistical approaches

The implementation of the statistical approaches applied in this paper are shown in Fig. 4. The development of lithofacies models for the Lothian Island site was carried out on through the multivariate and compositional data analyses of the each grain size bin along with multivariate statistical analyses of the log-normal and log skew-Laplace coefficients. For comparative purposes each model was compared using a plot of the first principal component scores following Andrews and Vogt (2014).

#### 3.4 CODA implementation

Centred log-ratio (clr) transformations were performed on the closed GSD prior to multivariate statistical analysis, with the transformations carried out using the 'compositions' package for the R statistical environment (R Development Core Team, 2010). R is freely available at <http://www.r-project.org> for Windows (current version 3.10.1, c. 26 MB). For the log-ratio transformation approach to be implemented, any zero-valued bins of quarter  $\phi$  intervals were removed. Channels of the grain-size distribution containing a zero in any of the observations were amalgamated and the arithmetic mean calculated, this was carried out on the 62.60  $\mu\text{m}$  to 2000  $\mu\text{m}$  fraction (i.e. 4.00  $\phi$  to -1.00  $\phi$ ). This amalgamation procedure generally leaves more than 80% of the channels unaltered (cf., Bloemsmas et al., 2012) and with a high level of redundancy in grain-size data (cf., Weltje and Prins, 2003).

### 3.5 Principal Components and Cluster Analysis of Sundarbans sediments

Principal components analysis, hierarchical cluster analysis and k-means cluster analysis were carried out on the log-transformed GSD from the Lothian Island core using the statistical package IBM® SPSS® Statistics version 19.0. Hierarchical cluster analysis using Ward's (1963) incremental sum of squares distances method based on squared Euclidean distances with the distance coefficients derived from the cluster analysis was used to determine the potential number of clusters. The k-means cluster analysis was subsequently performed in order to assign membership of samples to groups. These cluster analyses were carried out on the PCA scores, as well as on model coefficients from the log-normal and log skew-Laplace outputs.

### 3.6 Tests of stratigraphical randomness

A series of hypothesis tests was carried out in order to evaluate as to whether the stratigraphical grain size data series (e.g., principal component scores) were non-random, thereby, carrying a deterministic component of preferential organization (cf., Longhitano and Nemec, 2005). Applying tests of randomness allows for more robust estimation of trends in the lithofacies characteristics, as some tests may be more sensitive with regard to inherent data variability (Longhitano and Nemec, 2005). The non-parametric methods used are: the Wald-Wolfowitz runs tests, Mann-Kendall rank test, Cox-Stuart test, and Bartels ratio test; whereby the null hypothesis ( $H_0$ ) that the stratigraphical sequence data is random is evaluated (sig  $p < 0.05$ ). These statistical tests were carried out using the 'randtests' package for the R statistical environment (R Development Core Team, 2010).

### 3.7 Log-normal and log skew-Laplace distributions

Grain size distributions in the form of closed data from the Lothian Island core were examined through probability density functions (PDF) involving the log-normal and log skew-Laplace approaches. 'Gradistat' (Blott and Pye, 2001) was used to determine the log-normal distributions from the Lothian Island core through the product moment measurements (Krumbein and Pettijohn, 1938; Friedman and Johnson, 1982), using the four moment measurements of: mean, standard deviation, skewness, and kurtosis. The log skew-Laplace distributions from the Lothian Island GSD were determined using 'Shefsize' (Robson et al., 1997). This approach uses the maximum likelihood criterion along with a modified Davidson-Fletcher-Powell algorithm in order to derive the log-skew Laplace parameters of  $\alpha$  (corresponding to the fine grade coefficient),  $\beta$  (corresponding to the coarse grade coefficient), and  $\mu$  (the modal size) (see Walsh, 1975; Barndorff-Nielsen, 1977; Fieller et al., 1990;

Scalon et al., 2003). The model coefficients from the log-normal and log skew-Laplace distributions were examined using PCA and cluster analyses in order to produce lithofacies interpretations along with data-model and model intercomparisons.

## 4 Results

### 4.1 Lithofacies model generated using multivariate statistics of centre log-ratio transformed data

#### 4.1.1 PCA results from Lothian clr-transformed GSD data

Table 1 shows the correlation matrix of data used in the PCA of the clr-transformed grain size distributions from the Lothian Island core. Shown in bold there is a high degree of correlation (c.  $>0.70$ ) for several grain size fractions, in particular; fine, medium and coarse clay (c.  $12.02\phi - 8.97\phi$ ), very fine silt to coarse silt (c.  $8.00\phi - 6.01\phi$ ). Along with these, there is a particularly noticeable correlation between the very coarse silt to sand fraction (c.  $4.76\phi - 1.00\phi$ ), which also appear to be negatively correlated with the finer grain size distributions found. Thus, before PCA is carried out there appears to be two primary groups of material fine and coarse, that interact differently with intermediate size fractions (particularly silts). The PCA carried out on the data is summarised in Table 4, where the first principal component effectively accounts for 90% of the cumulative variance with an eigenvalue of 10.818. This is followed by second principal component comprising just over 6% of the remaining cumulative variance with an eigenvalue of 0.644.

As there was a distinctive break in variance explanation between the second and third principal components, only the first and second principal components were retained for further analysis as they represented the cumulative loading of 96% of the



overall variance of the dataset. The loadings (i.e., correlation coefficients between the original variable array and the principal component) from the first and second principal components are presented in Fig. 5. The positive coefficients show that the first principal component corresponds positively to size fractions from clay to coarse-silt, while relating negatively to very coarse-silt and sand fractions. The highest coefficients for the second principal component relate to the presence of medium silt and coarse grades of silt (the middle range of the sediment distribution), with slightly lower coefficients for the finer silt and clay. Sand and the very finest clays appear to show an inverse presence to the middle sediment range presence (negative coefficients on the second principal component).

Figure 6 shows the sample sediment score distributions for the Lothian Island core, in terms of each of the two new components extracted. For component 1 (Fig. 6A) shows, a general overall trend in reducing negative scores near the core base to c. 200 cm reflecting decreasing grain size, while the finest material (silts and clays) overlies this basal coarser material. The second principal component scores (Fig. 6B) show an overarching presence of fine to coarse silt throughout the core, albeit not on the same scale as the first principal component scores. These results indicate the Lothian Island sequence is composed almost entirely of silt with varying degrees of coarse silt and sand (at the base of the core) with increasing clay composition by the top of the sequence.

#### 4.1.2 Cluster analysis results from the Lothian Island clr-transformed GSD data

Hierarchical cluster analysis (HCA) and k-means cluster analysis indicate that four groups of sedimentary facies effectively explain the grain size distribution variation,

with their vertical disposition depicted in Fig. 7. The grain size association with each of the sedimentary facies:

- Facies 1 (f1): composed of clay with medium and fine silt;
- Facies 2 (f2): composed of medium and coarse silt with clay;
- Facies 3 (f3): composed of coarse silt with some sand, and;
- Facies 4 (f4): composed of sand (very fine sand to very coarse sand) with silt.

From the base of the core to a depth of 200 cm there are fluctuating trends in the facies; 4, 3, and 2. From 200 cm to the core surface, the cluster membership varies between facies 2 and 1. This fluctuating trend in cluster group association may reflect the PCA results. Facies; 4, 3, and 2 appear to show varying degrees of silt (coarse-medium-fine) and sand, while facies; 2, and 1 are characteristic of fine and medium silt with clay composition, with facies 1 composed mainly of clay. The vertical trend of these groups through the Lothian Island core indicates a fining-up sequence. Sand and silt fluctuations are indicated by f4 and f3, with f3 and f2 indicative of coarse silt to medium and fine silt with f1 representative of medium clay and fine-medium silt.

#### 4.1.3 Lithofacies model from Lothian Island clr-transformed GSD data

The association of these facies derived from the principal components analysis is further exemplified in Fig. 8, using a biplot of the first and second principal components, plus the facies association of the observations. The first principal component shows a fining-up trend (horizontal axis) whereas the second principal component depicts the overarching dominance of medium-silt with some coarse silt and sand. An examination of the PCA biplot quadrants reveals that samples located in the positive quadrant for PC1 and PC2 show medium silt and clay composition.

Samples present in the positive PC1 and negative PC2 quadrant are indicative of clay with medium silt. The dominance of f1 samples in these quadrants reflects this abundance of clay with medium silt in this facies. In the negative PC1 and PC2 quadrant samples are primarily sand (fine-medium-coarse), with f4 samples comprising this quadrant. This may be indicative of the sand composition near the base of the Lothian Island core. Along the negative PC1 and positive PC2 quadrant, the f3 samples are indicative of predominantly coarse-silt composition, with f2 samples characteristic of medium-silt with some coarse silt. The core stratigraphy indicates sedimentary facies transition order or potential stacking, with coarser silts plus sand at the base of the core, transitioning into an oscillating pattern of facies from c. 460 cm depth to c. 290 cm, which then passes into an almost static pattern of facies (2) from 290 cm until an almost completely homogeneous facies 1 from c. 200 cm to the core surface. Thus, the potential ordering of these sedimentary facies and their appearance in the core stratigraphy suggests that there are three distinctive broad stratigraphical Facies; (Fi) a lower coarser-silt with some sand; (Fii) medium-fine clay with fine-to-medium silt; and (Fiii) an upper fine-silt and clay, shown in Fig. 7.

## 4.2 Lithofacies models generated from the probability density function

### 4.2.1 PCA results from Lothian log-normal coefficients

The log-normal coefficients of the Lothian Island GSD indicated a fining-up trend in mean size along with samples being poorly sorted. GSDs were also found to be fine skewed and leptokurtic to mesokurtic throughout the core. Table 2 shows the correlation matrix of log-normal coefficients used in the PCA from the Lothian Island core. There is only one positive correlation present between skewness and

kurtosis (c. 0.33). These model coefficients tend to be negatively correlated with mean found to be most negatively correlated with skewness (c. -0.72). The PCA carried out on the data is summarised in Table 5, where the first principal component effectively accounts for c. 52% of the cumulative variance with an eigenvalue of 2.097. This is followed by second principal component comprising just over 29% of the remaining cumulative variance with an eigenvalue of 1.179.

There was a break in variance explanation between the second and third principal components, thus only the first and second principal components were retained for further analysis representing a cumulative loading of c. 82% of the overall variance. The loadings from the first and second principal components are presented in Fig. 9. The positive coefficients show that the first principal component corresponds positively to skewness and kurtosis, while relating negatively to mean size and sorting. The highest coefficients for the second principal component relate to sorting, with the other model coefficients possessing negative coefficient values.

Figure 10 shows the score distributions from the PCA of the log-normal coefficients. For component 1, positive scores characterise the overall trend from the core base to a depth of c. 350 cm indicative of skewness and kurtosis, while the mean and sorting variability overlies this trend from c. 350 cm to the core surface. The second principal component scores (Fig. 10b) appear to show an overarching presence of sorting variability from the core base to a depth of c. 200 cm; with mean, skewness, and kurtosis variability overlying this trend from c. 200 cm to the core surface. These results indicate the PCA of the log-normal coefficients from the Lothian Island sequence is characterised by variability driven by sorting, skewness, and kurtosis.

#### 4.2.2 Cluster analysis results from Lothian log-normal coefficients

Four groups of facies were derived from the HCA and k-means cluster analysis of the first and second principal component scores of the log-normal coefficients, with their vertical disposition depicted in Fig. 11. The grain size association with each of the sedimentary facies:

- Facies 1 (f1): samples reflecting sorting variability;
- Facies 2 (f2): samples indicative of skewness and kurtosis variability;
- Facies 3 (f3): samples indicative of mean size variability;
- Facies 4 (f4): samples most strongly indicative of sorting variability;

From the base of the core to a depth of c. 180 cm there are some fluctuations in facies but overwhelming the trend is driven by GSD facies 1. From 180 cm to the core surface, the cluster membership varies slightly between GSD facies 3 and 4. These fluctuating trends in GSD facies association may reflect the PCA results, with samples being primarily poorly sorted but decreasing in mean size.

#### 4.2.3 Lithofacies model from Lothian Island log-normal GSD data

The PCA biplot shown in Fig. 12, shows that samples located in the positive quadrant for PC1 and negative quadrant for PC2 show samples that may be considered to reflect skewness and kurtosis. In the negative PC1 and PC2 quadrant samples are primarily indicative of mean size variability and populated mostly with f3 samples. Along the negative PC1 and positive PC2 quadrant, the presence of f1 and f4 samples is indicative of sorting variability. Thus, the potential ordering of these sedimentary facies and their appearance in the core stratigraphy suggests that there are two distinctive broad stratigraphical facies; F(i)a, a sorting dominated

stratigraphic facies and F(ii)a, a sorting and mean size dominated stratigraphic facies, shown in Fig. 11.

#### 4.2.1 PCA results from Lothian log skew-Laplace coefficients

The log skew-Laplace coefficients of  $\alpha$ ,  $\beta$ ,  $\mu$ , and sorting ( $\alpha^2 + \beta^2$ ) developed by Olbricht (1982) (cf., Fieller et al., 1990) are used in the PCA and cluster analyses, respectively. The  $\alpha$  parameter values are considered to correspond to the fine grade coefficient, with  $\beta$  corresponding to the coarse grade coefficient, and  $\mu$  being the modal size (Fieller et al., 1992). The log skew-Laplace model revealed that  $\alpha$  appears to increase with a concomitant decrease in the  $\beta$  values moving up through the core, with  $\mu$  values found to be showing an overall decrease upwards. The overall trend appears to be one of very-well to moderately well-sorted GSD from the Lothian Island core, but with a general trend of decrease in grain size. The log skew-Laplace distributions appear to a general fining-up in grain size but being composed of fairly uniform size fractions. Table 3 shows the correlation matrix of log skew-Laplace coefficients used in the PCA. There is a strong positive correlation present between  $\alpha$  and sorting (c. 0.72) with a slightly lower positive correlation between  $\beta$  and sorting (c. 0.40). The model coefficients of  $\alpha$  and  $\beta$  along with  $\alpha$  and  $\mu$ , and  $\beta$  and  $\mu$  were found to be negatively correlated at c. -0.30, c. -0.18, and c. -0.18, respectively. The PCA carried out on the data is summarised in Table 6, where the first principal component comprises c. 47% of the cumulative variance with an eigenvalue of 1.898. The second principal component comprises just over 32% of the remaining cumulative variance with an eigenvalue of 1.284.

The first and second principal components were retained for further analysis comprising a cumulative loading of c. 80% of the variance. The loadings from the

first and second principal components are presented in Fig. 13. The positive coefficients show that the first principal component correspond positively to  $\alpha$ ,  $\beta$ , and sorting, while relating negatively to  $\mu$ . The highest coefficients for the second principal component relate to  $\beta$  followed by sorting, with the other model coefficients possessing negative coefficient values.

Figure 14 shows the score distributions from the PCA of the log skew-Laplace coefficients. For component 1, negative scores characterise the overall trend from the core base to a depth of c. 190 cm indicative of  $\mu$ , while  $\alpha$ ,  $\beta$ , and sorting variability overlies this trend from c. 190 cm to the core surface. The second principal component scores (Fig. 14b) appears to show a presence of  $\alpha$  and  $\mu$  variability with some samples through the sequence reflecting  $\beta$  and sorting variability. These results indicate the PCA of the log skew-Laplace coefficients from the Lothian Island sequence is characterised by variability driven by modal size and the variability between fine and coarse sediment deposition relative to the sorting taking place.

#### 4.2.2 Cluster analysis results from Lothian log skew-Laplace coefficients

Three groups of facies were derived from the HCA and k-means cluster analysis of the first and second principal component scores of the log skew-Laplace coefficients, with their vertical disposition shown in Fig. 15. The grain size association with each of the sedimentary facies:

- Facies 1 (f1): samples reflecting sorting variability;
- Facies 2 (f2): samples indicative of  $\alpha$  and  $\mu$  variability;
- Facies 3 (f3): samples indicative of  $\beta$  and sorting variability;

From the base of the core to a depth of c. 380 cm there are fluctuations between facies 3 and 2. From c. 380 cm to c. 190 cm, the samples consist of mainly f3. Finally, from c. 190 cm to the core surface the samples are composed of f2 with some f1 samples.

#### 4.2.3 Lithofacies model from Lothian Island log skew-Laplace GSD data

The PCA biplot quadrants shown in Fig. 16 reveals that samples located in the positive quadrant for both the first and second principal components may be understood to reflect  $\beta$  and sorting variability, and composed of some f3 and all f1 samples. In the positive first principal components and negative second principal components quadrant, samples are indicative of  $\alpha$  variability and populated mostly with f2 samples. Along the negative second and first principal components quadrant,  $\mu$  variability appears to underlie this region with the presence of f2 and f3 samples. A large proportion of f3 samples appear to be present in the negative first principal components and positive second principal components quadrant. The sedimentary facies that may be determined from this analysis suggests three broad stratigraphical facies; with F(i)b, consisting of variability present with all of the log skew-Laplace coefficients that is overlain with F(ii)b, a predominantly coarser and sorting dominated stratigraphic facies. These stratigraphic facies are finally overlain with and finer and modal size dominated stratigraphic facies F(iii) b, shown in Fig. 15.

#### 4.3 Tests of stratigraphical randomness from the lithofacies models

A variety of randomness tests on the Lothian GSD (clr-transformed data) show that for the first principal component scores and cluster groups shown in Table 7, there are significant, non-random trends present ( $p < 0.001$ ). These randomness tests on the



first principal component scores and cluster groups from the log-normal and log skew-Laplace coefficients also show a non-random trend present with  $p < 0.001$ .

4.4 Model inter-comparison: efficacy of statistical analyses of grain size distributions

Model intercomparisons through plotting of the first principal components scores are shown in Fig. 17 for the log-normal and clr-transformed data. A least squares regression shows a poor, negative correlation coefficient ( $R^2$ ) value of 0.328 (Fig. 17A). A similar  $R^2$  value at 0.354 was found with the log skew-Laplace first component scores against the clr-transformed scores (Fig. 17B). Plotting the log-skew Laplace first component scores against those of the log-normal scores also revealed a very poor correlation between the data with an  $R^2$  value at 0.023 (Fig. 17C). There appears to be very poor correlation between the probability density function models (i.e., first component scores from log-normal and log skew-Laplace coefficients) and the clr-transformed data first component scores. Furthermore, the PDF first component scores illustrate a very poor correlation between the log-normal and log skew-Laplace models.

## 5. Discussion

### 5.1 Efficacy of lithofacies determination using multivariate and compositional data analysis approaches

The primary proposition of this paper was to illustrate the utility of multivariate statistics and compositional data analysis in the interpretation of GSD facies alongside qualitative descriptions of GSD and multivariate statistical analyses of PDF approaches. The key point is that there are inherent inconsistencies between the distribution approaches (e.g., the calculation for log-normal is different from that of

the log skew-Laplace), along with the least squares regression of the first principal component scores, shown in the model intercomparison (section 4.4).

The advantages of employing a compositional data analysis approach coupled with multivariate statistics (PCA and cluster analysis) is that the entire GSD is considered in the analysis, which is not the case with PDF approaches. Deltaic sedimentation is a highly dynamic process composed of auto- and allocyclic variability; with GSDs deposited under highly fluctuating conditions, which in the case of the Sundarbans are dictated by a mixture of extreme surge events, seasonal monsoonal and semi-diurnal tidal variability (Flood, 2014). Capturing this variability by decomposing the dataset into dimensionally comprehensible components allows for the retention of maximum variance through the PCA scheme (e.g., c. 96% of Lothian Island GSD variance characterised by two principal components). Applying such dimensionality reduction to the PDF log-normal and log skew-Laplace coefficients is somewhat redundant as the dataset has already been decomposed into a series of model coefficients, where data have been masked through the model fit (cf., Roberson and Weltje, 2014). PCA applied to the log-normal coefficients in this study found that c. 52% and 29% of the variance was characterised by the first and second principal components (shown in Table 5). Similarly, applying PCA to the log skew-Laplace  $\alpha$ ,  $\beta$ ,  $\mu$ , and  $\alpha^2 + \beta^2$  (i.e., sorting analogue) found that the first and second principal component variances were c. 47% and 32%, respectively (shown in Table 6). The reduction of dimensionality in a PDF approach is at the expense of overall data variance, which can be avoided when employing the whole GSD. Interpretation of the score plots from the log-normal coefficients (Figs. 10, 12) and the log skew-Laplace coefficients PCA (Figs. 14, 16) using the respective loadings (Figs. 9, 13) is somewhat convoluted. Scores and loadings from the application of the

PCA to the model coefficients reveal only the variability of the model coefficients as opposed to actual grain size variance; which characterises the PCA of the clr-transformed data analysis.

Cluster analysis of the PCA scores from the PDF coefficients usually leads to a more agglomerated textural and stratigraphic facies statement. This is illustrated in Figs. 11 and 15, with the former illustrating vertical disposition of GSD facies from the log-normal and log skew-Laplace. Although facies tend to fluctuate, there is an apparent agglomeration of samples into uniform facies when compared to that of the CODA based facies (Fig. 7). The derivation of facies based on the cluster analyses of the PCA scores from the model coefficients shows that in the case of log-normal there are up to four facies identified, with log skew-Laplace producing only three facies. The robustness therefore of textural-based facies and that of broader stratigraphic facies is questionable, as: (1) are facies reliably distinctive in either (e.g., how 'facies' 1, 2, 3, and 4 in the log-normal scheme adequately differ), and; (2) does the agglomeration of the GSD based on PDF coefficients homogenise the facies variability (e.g., as exemplified with the log skew-Laplace 'facies'). However, the crux of these arguments may lie in the use of model coefficients that decompose data in a reductionist manner with the loss of inherent variance. The application of tests of randomness to the PCA scores and GSD facies of the log-normal (shown in Figs. 10, 11) and log skew-Laplace (shown in Figs. 14, 15) coefficients reveal an apparent non-random trend (shown in Table 4). Although there is a non-random trend present in the decomposition of these PDFs into scores and GSD facies, these trends however may reflect statistically derived variability.

The key challenge posed by using PDFs is that there is no sedimentological argument for the use of such model coefficient parameters as criteria for judging the

efficacy, as such models are premised on using non-geological criteria (e.g., goodness-of-fit) (Weltje and Prins, 2007). Furthermore, interpretation of grain size facies and broad sedimentary facies from the PDF approaches appears to be somewhat difficult as the interpretation is based on the relative variance accounted for by model coefficients and not on the grain size fractions. Furthermore, the retention of variability prior to the application of dimensionality reduction is unknown as the parametric models used are dependent upon the investigator outlining the number of distributions and the parameters used (Weltje and Prins, 2007). Thus, although the PCA of the log-normal and log skew-Laplace coefficients presented in section 4.2, characterises upwards of 80% of the variance accounted; the empirical data variance prior to model fitting is unknown and may be reduced as a result (Roberson and Weltje, 2014).

## 5.2 Lithofacies models: intercomparison of approaches

Data model comparisons shown in Fig. 17, show that there is only a very slight correlation between the first component scores from the PDFs with those of the log-transformed dataset (Fig. 17A, Fig. 17B, c.  $R^2 = 0.33$  and  $0.35$ ), with a very poor correlation found between the first component scores of each of the PDFs (shown in Fig. 17C,  $R^2 = 0.023$ ). There appears to be some level of inconsistency in terms of the models but most notably between the two PDF approaches used. Both PDF approaches to grain size distributions seek to capture grain size variability in a continuous function with as few parameters as possible (Weltje and Prins, 2007). Although these parameters are largely fewer by comparison to employing the whole GSD, the *a priori* assumption when comparing the scores of the log-transformed data, log-normal coefficients, and log skew-Laplace coefficients is that they should

capture the overall variability in the dataset. The lack of any consistency, especially between the parametric model scores appears to demonstrate problems of using such parametric models in that they cannot be fully compared to one another. It can be interpreted therefore that the PDF approaches are variable or misaligned with one another. The difference in nature of these approaches however may be best illustrated in Fig. 18, where the log-normal distributions of the Lothian Island grain size data are depicted. The sample distributions are primarily bimodal and there are only very slight differences between facies shown. Therefore there is some arbitrariness as to the assessment of these models as being statistical representations of the original grain size distributions.

### 5.3 Comparison of proposed lithofacies with existing stratigraphic facies for the Sundarbans

Three broader stratigraphic facies have been identified in Lothian Island core data by comparison to Allison et al. (2003) facies classification. At the outset, it should be noted that contextually and methodologically; a direct comparison of broad stratigraphic facies between this study and that of Allison et al. (2003) is moot given the spatial variability of cores and the overall broader scheme of Allison et al. (2003) facies succession for the entirety of the lower G-B delta during the Holocene epoch. However, the statistically derived facies presented in this paper can be considered to range over two of the principal facies that Allison et al. (2003) found in their study (i.e., that of the 'intertidal shoal' and 'supratidal' facies). Therefore, although these studies are comparatively different in execution, they are not however mutually exclusive. As this study has shown that the sedimentary facies derived in a statistically robust manner may complement the generalised facies scheme of Allison

et al. (2003). In terms of the statistical analyses of grain size distributions, the use of multivariate statistical approaches is not novel when applied against the backdrop of existing studies (e.g., Davis, 1970; Chambers and Upchurch, 1979; Lirer and Vinci, 1991; Zhou et al., 1991; Weltje and Prins, 2003; Weltje and Prins, 2007), nor even studies on the Sundarbans (cf., Purkait and Majumdar (2014)). However, formalising a schema of approaches founded in compositional data analysis offers an unequivocal solution to dealing with polymodal, non-unique PDFs (Roberson and Weltje, 2014). In this regard, this study may be seen as a logical progression from the statistical approaches of analysis of grain size distributions from that of Purkait and Majumdar (2014).

#### 5.4 Potential limitations of multivariate approaches presented

The limitations of the facies analyses approaches advocated in this paper is that of the sampling regime pursued. GSD data was derived from an 8cm sampling interval, which in the case of deriving a robust facies interpretation leads to some issues regarding windowing or aliasing. Sampling should generally incorporate as broad a range of samples and potential GSDs as possible. However, if the sampling regime is too low, the opportunity to derive these GSDs is limited and the development of a facies interpretation is not as robust. The questions then arise as to whether there are ‘real’ differences in GSD compositions between facies: how different are facies 1 and 2 compared to facies 2 and 3, and whether facies 2 and 3 are actually different to facies 3 and 4. In order to resolve such issues, higher density of sampling may be required. However, the transition between samples showing non-randomness may also be considered indicative of some deterministic GSD trend which is not solely a function of sampling interval.

## 6. Conclusions

This paper represents a framework of methodology for the analysis of grain size distributions through compositional data analysis and multivariate statistics. The statistical methodology outlined here consists of employing a centred log-ratio transformation on a closed GSD dataset. This is followed by eigenvalue and cluster analysis by which facies can be identified. The methodology is based on limiting the subjectivity of analysis of grain size distributions that generally characterizes probability density function approaches. The CODA methodology removes challenges posed by constrained data with PCA extracting the maximum variance present, with this variance examined categorically through a series of cluster analyses. The primary conclusions from this study are:

- (i) The CODA approach allows data to be examined in an unconstrained environment;
- (ii) With PCA, such log-transformed GSD data may be characterised through dimensionality reduction into a 'simpler' (i.e. interpretable) model explaining the maximised variance of the data;
- (iii) Cluster analysis techniques allow for an agglomeration of similar data groups that may be interpreted as indicative of varying environmental conditions;
- (iv) Probability density function approaches (i.e., log-normal and log skew-Laplace) may offer statistically tractable estimations of grain size variability, they do however obfuscate interpretations of grain size variance and pose inherent difficulties where polymodality is present in the GSD;
- (v) These drawbacks of PDF approaches are overcome through the use of CODA and employing the entire GSD through the multivariate approaches outlined in this paper, with the advantages of objectively characterising the sedimentary facies.

**Acknowledgements**

RF acknowledges both the supported provided by a Research Studentship from the Department for Employment and Learning (Northern Ireland), and the support of the Department of Education and Science's Higher Education Grant Scheme (ROI) provided through Laois County Council (ROI). RF also acknowledges the School of Geography, Archaeology and Palaeoecology, Queen's University, Belfast for the fieldwork support provided by their Soulby Research Fund. The authors would like to acknowledge the editorial comments from Jasper Knight and Gert Jan Weltje which were found to be very helpful in the modification of the manuscript.



## References

- Aitchison, J., 1982. The statistical analysis of compositional data. *Journal of the Royal Statistical Society Series B (Methodological)* 44, 139–177.
- Aitchison, J., 1986. The statistical analysis of compositional data: Monographs on statistics and applied probability. Chapman & Hall Ltd., London. 436 pp.
- Allen, G.P., Castaing, P., Klingebiel, A., 1971. Preliminary investigation of the surficial sediments in the Cap-Breton Canyon (southwest France) and the surrounding continental shelf. *Marine Geology* 10, M27–M32.
- Allen, G.P., Castaing, P., Klingebiel, A., 1972. Distinction of elementary sand populations in the Gironde estuary (France) by R-mode factor analysis of grain-size data. *Sedimentology* 19, 21–35.
- Allison, M.A., Khan, S.R., Goodbred, Jr., S.L., Kuehl, S.A., 2003. Stratigraphic evolution of the late Holocene Ganges–Brahmaputra lower delta plain. *Sedimentary Geology* 155, 317–342.
- Andrews, J.T., Vogt, C., 2014. Source to sink: Statistical identification of regional variations in the mineralogy of surface sediments in the western Nordic Seas (58°N–75°N; 10°W–40°W). *Marine Geology* 357, 151–162.
- Bagnold, R., Barndorff-Nielsen, O., 1980. The pattern of natural size distributions. *Sedimentology* 27, 199–207.
- Barndorff-Nielsen, O., 1977. Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London A (Mathematical and Physical Sciences)* 353, 401–419.
- Beierle, B.D., Lamoureux, S.F., Cockburn, J.M., Spooner, I., 2002. A new method for visualizing sediment particle size distributions. *Journal of Paleolimnology* 27, 279–283.

- Bloemsma, M.R., 2010. Semi-Automatic Core Characterisation based on Geochemical Logging Data, Unpublished M.Sc. Thesis: Delft University of Technology, 144 pp.
- Bloemsma, M.R., Zabel, M., Stuut, J.B.W., Tjallingii, R., Collins, J.A., Weltje, G.J., 2012. Modelling the joint variability of grain size and chemical composition in sediments. *Sedimentary Geology* 280, 135–148.
- Blott, S.J., Pye, K., 2001. GRADISTAT: a grain size distribution and statistics package for the analysis of unconsolidated sediments. *Earth Surface Processes and Landforms* 26, 1237–1248.
- Chambers, R.L., Upchurch, S.B., 1979. Multivariate analysis of sedimentary environments using grain-size frequency distributions. *Mathematical Geology* 11, 27–43.
- Chayes, F., 1960. On correlation between variables of constant sum. *Journal of Geophysical Research* 6, 4185–4193.
- Chayes, F., 1971. Ratio correlation: a manual for students of petrology and geochemistry. University of Chicago Press, Chicago, 99 pp.
- Dal Cin, R., 1976. The use of factor analysis in determining beach erosion and accretion from grain-size data, *Marine Geology* 20, 95–116.
- Davis, J.C., 1970. Information contained in sediment-size analyses. *Mathematical Geology* 2, 105–112.
- Davis, J.C., 1986. *Statistics and Data Analysis in Geology*, 2nd edn. John Wiley and Sons, New York, 646 pp.
- East, T.J., 1985. A factor analytic approach to the identification of geomorphic processes from soil particle size characteristics. *Earth Surface Processes and Landforms* 10, 441–463.

- East, T.J., 1987. A multivariate analysis of the particle size characteristics of regolith in a catchment on the Darling Downs, Australia. *Catena* 14, 101–118.
- Fieller, N.R.J., Flenley, E.C., Gilbertson, D.D., Thomas, D.S.G., 1990. Dumb-bells: a plotting convention for “mixed” grain size populations. *Sedimentary Geology* 69, 7–12.
- Fieller, N., Flenley, E., Olbricht, W., 1992. Statistics of particle size data. *Journal of the Royal Statistical Society Series C (Applied Statistics)* 41, 127–146.
- Fieller, N.R.J., Flenley, E.C., 1992. Statistics of particle size data. *Journal of Applied Statistics* 41, 127–146.
- Fieller, N.R.J., Gilbertson, D.D., Olbricht, W., 1984. A new method for environmental analysis of particle size distribution data from shoreline sediments. *Nature* 311, 648–651.
- Flood, R.P., 2014. Post Mid-Holocene Sedimentation of the West Bengal Sundarbans. Unpublished Ph.D. Thesis, Queen’s University, Belfast, 646 pp.
- Folk, R.L., Ward, W.C., 1957. Brazos River bar: a study in the significance of grain size parameters. *Journal of Sedimentary Petrology* 27, 3–26.
- Forrest, J., Clark, N.R., 1989. Characterizing grain size distributions: evaluation of a new approach using multivariate extension of entropy analysis. *Sedimentology* 36, 711–722.
- Fredlund, M., Fredlund, D., Wilson, G., 2000. An equation to represent grain size distribution. *Canadian Geotechnical Journal* 37, 817–827.
- Friedman, G.M., Johnson, K.G., 1982. *Exercises in Sedimentology*. John Wiley and Sons, New York, 208 pp.
- Friedman, G.M., 1962. On sorting, sorting coefficients, and the lognormality of the grain-size distribution of sandstones. *The Journal of Geology* 70, 737–753.

- Jain, A.K., 2010. Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* 31, 651–666.
- Goodbred Jr., S.L., Kuehl, S.A., 2000. The significance of large sediment supply, active tectonism, and eustasy on margin sequence development: late Quaternary stratigraphy and evolution of the Ganges-Brahmaputra delta. *Sedimentary Geology* 133, 227–248.
- Jöreskog, K.G., Klován, J.E., Reyment, R.A., 1976. *Geological Factor Analysis*. Elsevier, Amsterdam, 178 pp.
- Kasper-Zubillaga, J.J., Carranza-Edwards, A., 2005. Grain size discrimination between sands of desert and coastal dunes from northwestern Mexico. *Revista Mexicana de Ciencias Geológicas* 22, 383–390.
- Klován, J.E., 1966. The use of factor analysis in determining environments from grain-size distributions. *Journal of Sedimentary Petrology* 36, 57–69.
- Krumbein, W.C., 1934. Size frequency distributions of sediments. *Journal of Sedimentary Petrology* 4, 65–77.
- Krumbein, W.C., 1938. Size frequency distribution of sediments and the normal phi curve. *Journal of Sedimentary Petrology* 8, 84–90.
- Lirer, L., Vinci, A., 1991. Grain-size distributions of pyroclastic deposits. *Sedimentology* 38, 1075–1083.
- Longhitano, S.G., Nemec, W., 2005. Statistical analysis of bed-thickness variation in a Tortonian succession of biocalcarenitic tidal dunes, Amantea Basin, Calabria, southern Italy. *Sedimentary Geology* 179, 195–224.
- Moiola, R.J., Spencer, A.B., 1979. Differentiation of eolian deposits by discriminant analysis. In: McKee, E.D. (Ed.), *A Study of Global Sand Seas: USGS Professional Papers*, vol. 1052, pp. 53–58.

- Olbricht, W., 1982. Modern statistical analysis of ancient sands. M.Sc. dissertation, University of Sheffield, 86 pp.
- Passega, R., 1957. Texture as characteristic of clastic deposition. American Association of Petroleum Geology Bulletin 4, 1952–1984.
- Passega, R., 1964. Grain size representation by CM patterns as a geological tool. Journal of Sedimentary Petrology 34, 830–847.
- Pawlowsky-Glahn, V., Egozcue, J.J., 2006. Compositional data and their analysis: an introduction. In: Buccianti, A., Mateu-Figueras, G., Pawlowsky-Glahn, V. (Eds.), Compositional Data Analysis in the Geosciences – From Theory to Practice. Geological Society of London, Special Publication 264, pp. 1–10.
- Pearson, K., 1896. On the form of spurious correlations which may arise when indices are used in the measurement of organs. Proceedings of the Royal Society, London 60, 489–502.
- Purkait, B., Majumdar, D.D., 2014. Distinguishing different sedimentary facies in a deltaic system. Sedimentary Geology 308, 53–62.
- Rencher, A.C., 2002. Methods of Multivariate Analysis. John Wiley and Sons, New Jersey, 738 pp.
- Roberson, S., Weltje, G.J., 2014. Inter-instrument comparison of particle-size analysers. Sedimentology 61, pp. 1157–1174.
- Robson, D.R., Fieller, N.R.J., Stillman, E.C., 1997. SHEFSIZE: A Windows Program for Analysing Size Distributions. School of Mathematics and Statistics, University of Sheffield.
- Sahu, B.K., 1964. Depositional mechanisms from the size analysis of clastic sediments. Journal of Sedimentary Petrology 34, 73–84.

- Sarnthein, M., Tetzlaff, G., Koopmann, B., Wolter, K., Pflaumann, U., 1981. Glacial and interglacial wind regimes over the eastern subtropical Atlantic and North-West Africa. *Nature* 293, 193–196.
- Scalon, J., Fieller, N.R.J., Stillman, E.C., Atkinson, H.V., 2003. A model-based analysis of particle size distributions in composite materials. *Acta Materialia* 51, 997–1006.
- Sneath, P.H.A., Sokal, R.R., 1973. Numerical taxonomy the principles and practice of numerical classification. Freeman, San Francisco, 588 pp.
- Solohub, J.T., Klován, J.E., 1970. Evaluation of grain-size parameters in lacustrine environments. *Journal of Sedimentary Petrology* 40, 81–101.
- Stauble, D.K., Cialone, M.A., 1996. Sediment dynamics and profile interactions: Duck94. *Coastal Engineering Proceedings*, 1(25), 3921–3934.
- Swan, A.R.H., Sandilands, M.H., 1995. Introduction to geological data analysis. Blackwell Science, Oxford, 464 pp.
- Syvitski, J., 1991. Principles, Methods, and Application of Particle Size Analysis. Cambridge University Press, New York, 368 pp.
- Templ, M., Filzmoser, P., Reimann, C., 2008. Cluster analysis applied to regional geochemical data: Problems and possibilities. *Applied Geochemistry* 23, 2198–2213.
- Van den Boogaart, K.G., Tolosana-Delgado, R., 2013. Analyzing compositional data with R. Springer, Berlin, 258 pp.
- Walsh, G.R., 1975. Methods for optimization. John Wiley and Sons, New York, 200 pp.
- Ward, J.H., 1963. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* 58, 236–244.

- Weltje, G.J., Prins, M.A., 2003. Muddled or mixed? Inferring palaeoclimate from size distributions of deep-sea clastics. *Sedimentary Geology* 162, 39–62.
- Weltje, G.J., Prins, M.A., 2007. Genetically meaningful decomposition of grain-size distributions. *Sedimentary Geology* 202, 409–424.
- Weltje, G.J., Tjallingii, R., 2008. Calibration of XRF core scanners for quantitative geochemical logging of sediment cores: theory and application. *Earth and Planetary Science Letters* 274, 423–438.
- Wishart, D., 1969. An Algorithm for Hierarchical Classifications. *Biometrics* 25, 165–170.
- Zhou, D., Chen, H., Lou, Y., 1991. The log-ratio approach to the classification of modern sediments and sedimentary environments in northern South China Sea. *Mathematical Geology* 23, 157–165.

## **FIGURE AND TABLE CAPTIONS**

### **Figure captions**

Fig. 1. Extent of the tidal delta complex, (a) West Bengal Sundarbans (India) and (b) East Bengal Sundarbans (Bangladesh), within the tidal delta plain (adapted from Rogers et al., 2013).

Fig. 2. Sites cored in the West Bengal Sundarbans, India (November 2010).

Fig. 3. Series of semi-log plots of different simulated grain size frequency data with each subplot containing a unimodal, bimodal, trimodal and quadramodal frequency distribution possessing identical log-normal mean and standard deviation values (after Roberson and Weltje, 2014).

Fig. 4. Flow diagram of model implementation and data analysis pursued in this study.

Fig. 5. First (a) and second (b) principal component loadings for each grain size class of the clr-transformed Lothian Island GSD.

Fig. 6. PCA score plots for first (a) and second (b) principal components of the clr-transformed Lothian Island core GSD.

Fig. 7. Vertical disposition of the broad stratigraphical facies throughout the clr-transformed Lothian Island GSD core data. These broader stratigraphical facies based on the sedimentological facies consist of: (i) lower coarser-silt with some sand; (ii) middle medium-fine clay with fine-to-medium silt; and (iii) upper fine-silt and clay.

Fig. 8. Biplot of first (a) and second (b) principal components with cluster group association of clr-transformed Lothian Island core GSD.

Fig. 9. First (a) and second (b) principal component loadings for each grain size class of each log-normal coefficient of the Lothian Island GSD.

Fig. 10. PCA score plots for first (a) and second (b) principal components of the log-normal coefficients.

Fig. 11. Vertical disposition of the broad stratigraphical facies throughout the Lothian Island core derived from the log-normal model coefficients. These broader stratigraphical facies consist of: F(i)a, a sorting dominated stratigraphic facies and F(ii)a, a sorting and mean size dominated stratigraphic facies.

Fig. 12. Biplot of first (a) and second (b) principal components with cluster group association of log-normal Lothian Island core GSD.

Fig. 13. First (a) and second (b) principal component loadings for each grain size class of each skew-Laplace coefficient coefficient of the Lothian Island GSD.



Fig. 14. PCA score plots for first (a) and second (b) principal components of the log skew-Laplace coefficients of the Lothian Island GSD.

Fig. 15. Vertical disposition of the broad stratigraphical facies throughout the Lothian Island core derived from the log skew-Laplace model coefficients. These broader stratigraphical facies consist of: F(i)b, variability present in all log skew-Laplace coefficients; F(ii)b, coarse and sorting dominated stratigraphic facies, and; F(iii)b, fine and modal size dominated stratigraphic facies.

Fig. 16. Biplot of first (a) and second (b) principal components with cluster group association of log skew-Laplace Lothian Island core GSD.

Fig. 17. Least squares regression of (A) 1<sup>st</sup> principal component (PC) scores of the log-normal coefficients and (B) log skew-Laplace coefficients against the 1<sup>st</sup> PC scores of the clr-transformed data with a regression of the (C) 1<sup>st</sup> PC scores of the log skew-Laplace coefficients against those 1<sup>st</sup> PC scores of the log-normal coefficients.

Fig. 18. Log-normal distributions from the Lothian Island core with the clr-transformed cluster facies highlighted.

**Tables**

Table 1. Correlation matrix of GSD bins for Lothian Island core.

Table 2. Correlation matrix of log-normal coefficients for Lothian Island core.

Table 3. Correlation matrix of log skew-Laplace coefficients for Lothian Island core.

Table 4. Total variance explained by the PCA on the clr-transformed data from Lothian Island core.

Table 5. Total variance explained by the PCA on the log-normal coefficients from Lothian Island core.

Table 6. Total variance explained by the PCA on the log skew-Laplace coefficients from Lothian Island core.

Table 7. Post-hoc statistical tests of randomness on the first principal component scores and GSD facies (clusters) from the clr-transformed data, log-normal, and log skew-Laplace coefficients from the Lothian Island core.

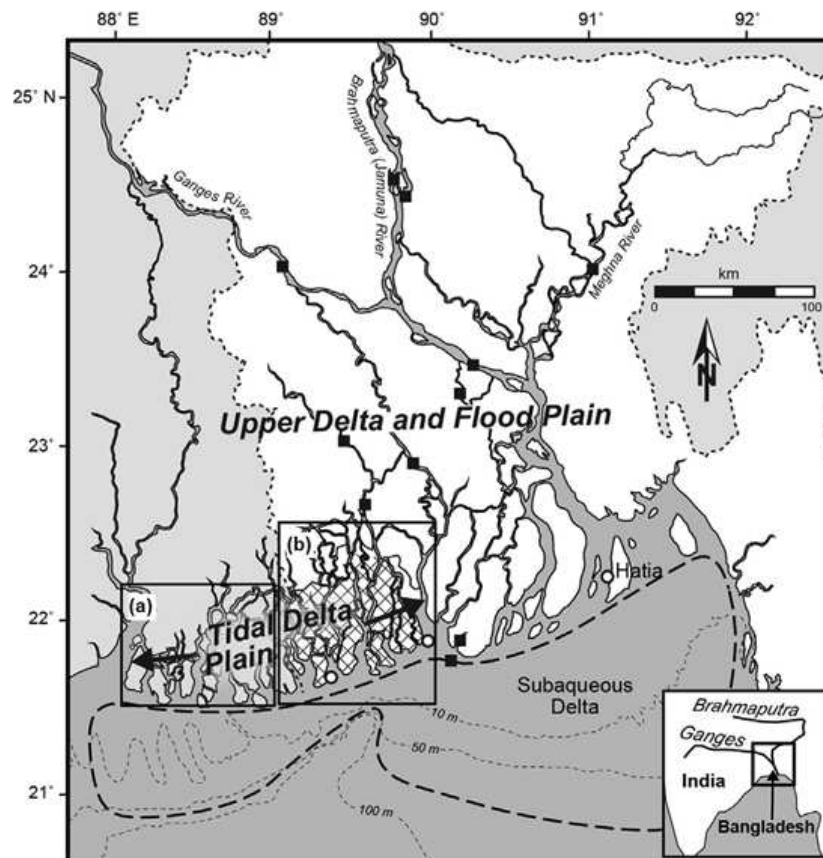


Figure 1

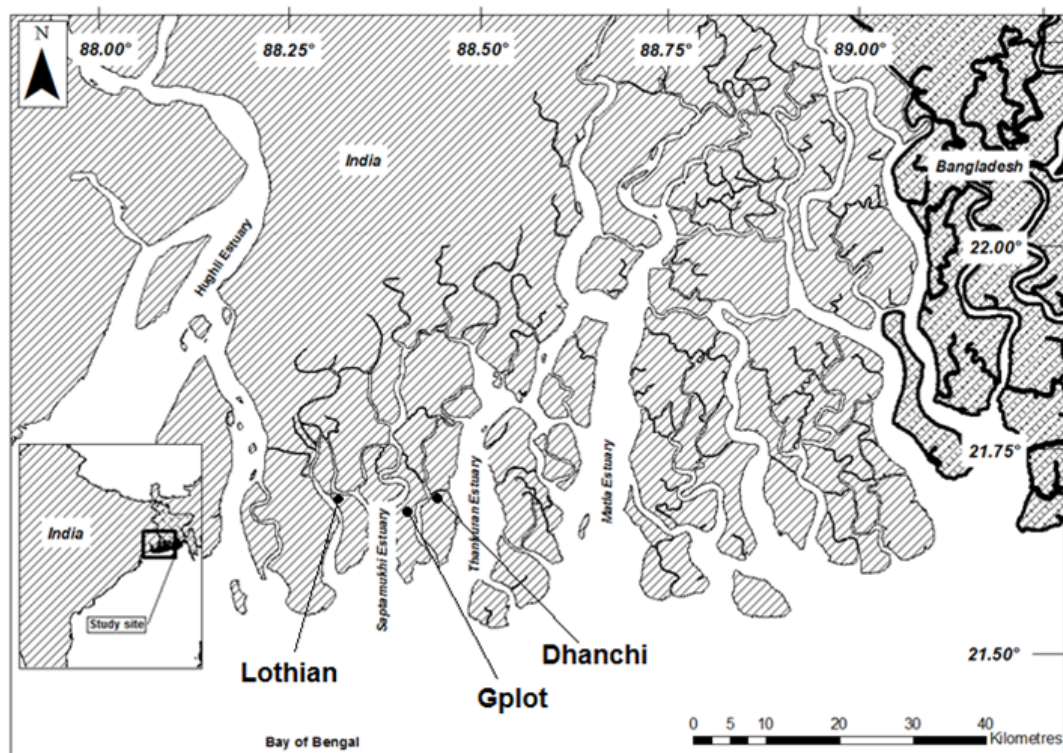
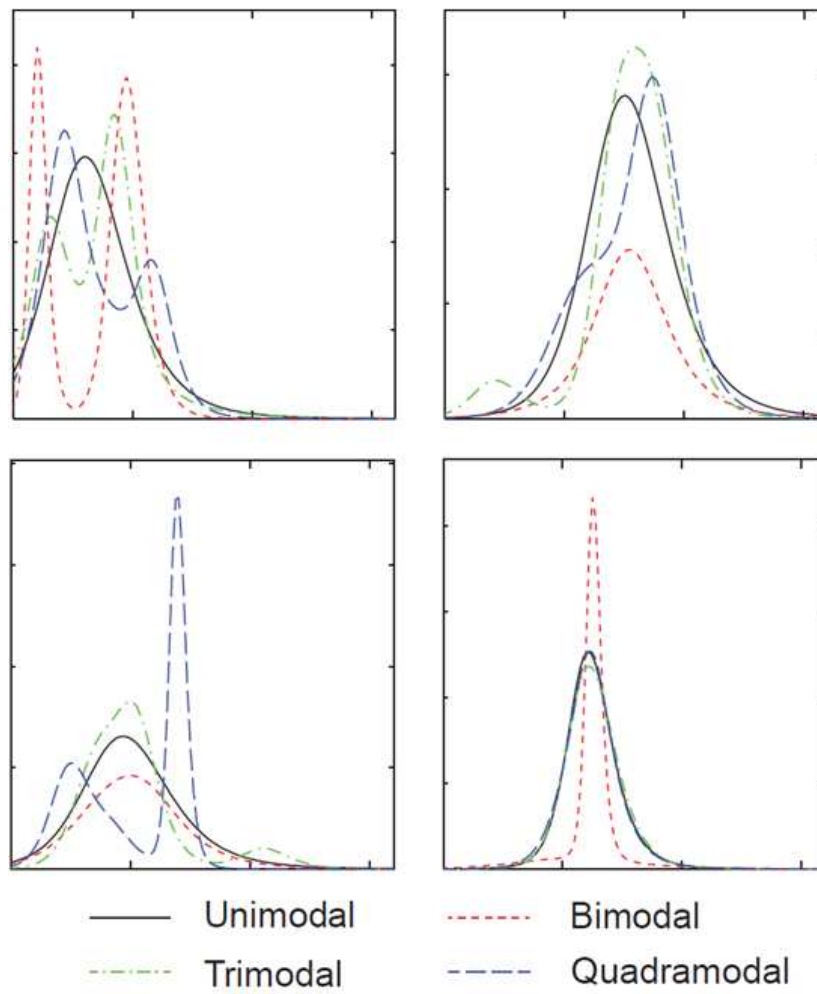


Figure 2

**Figure 3**

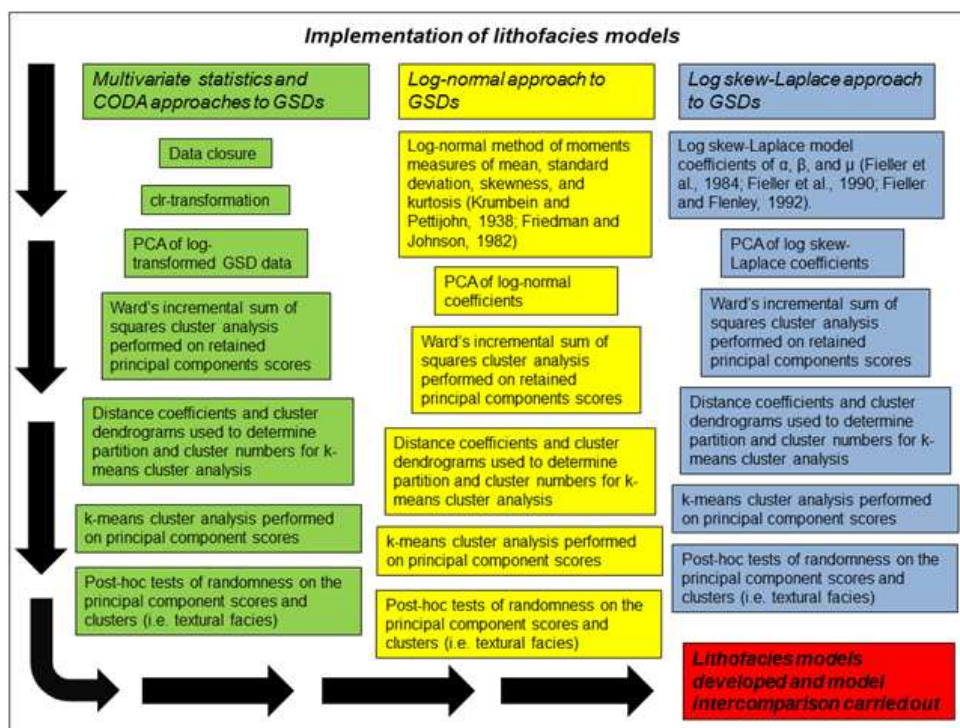


Figure 4

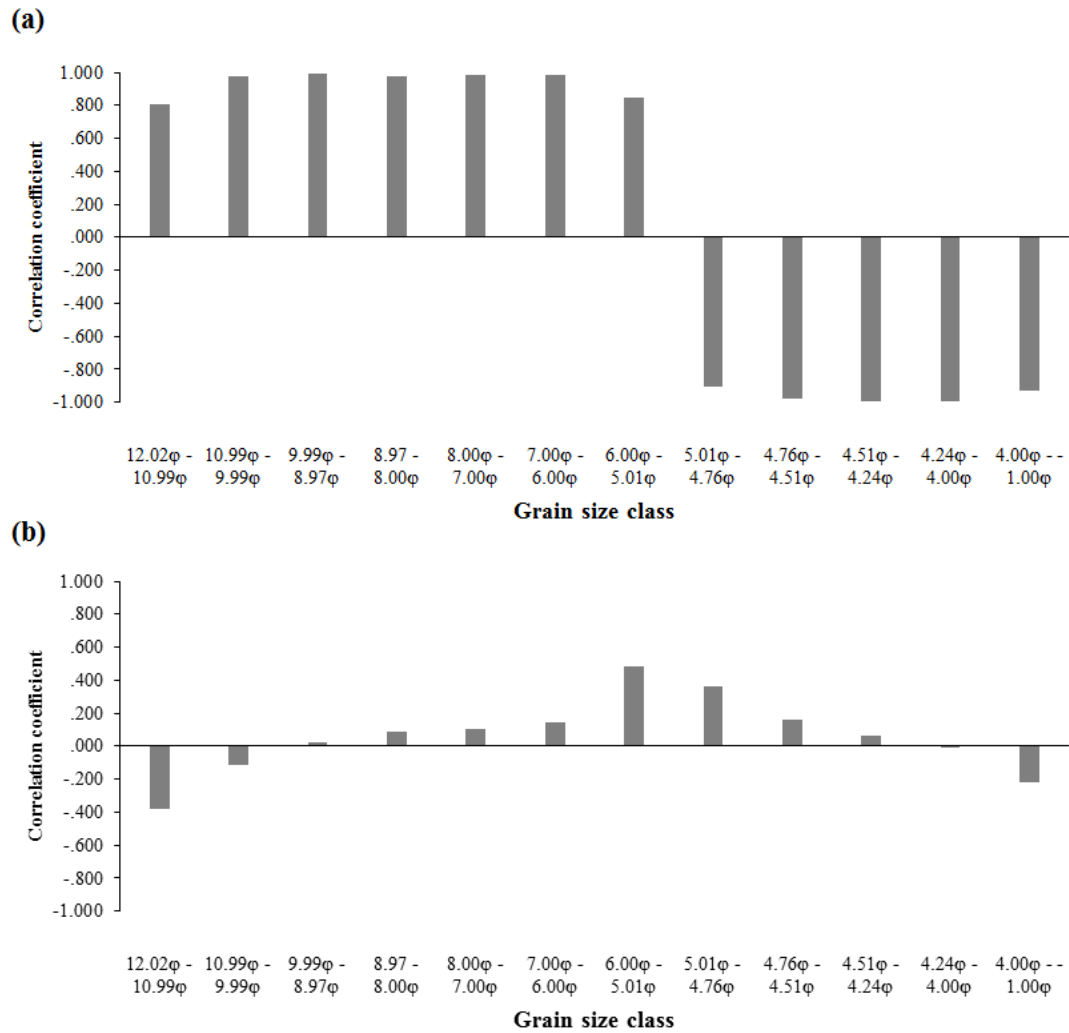


Figure 5

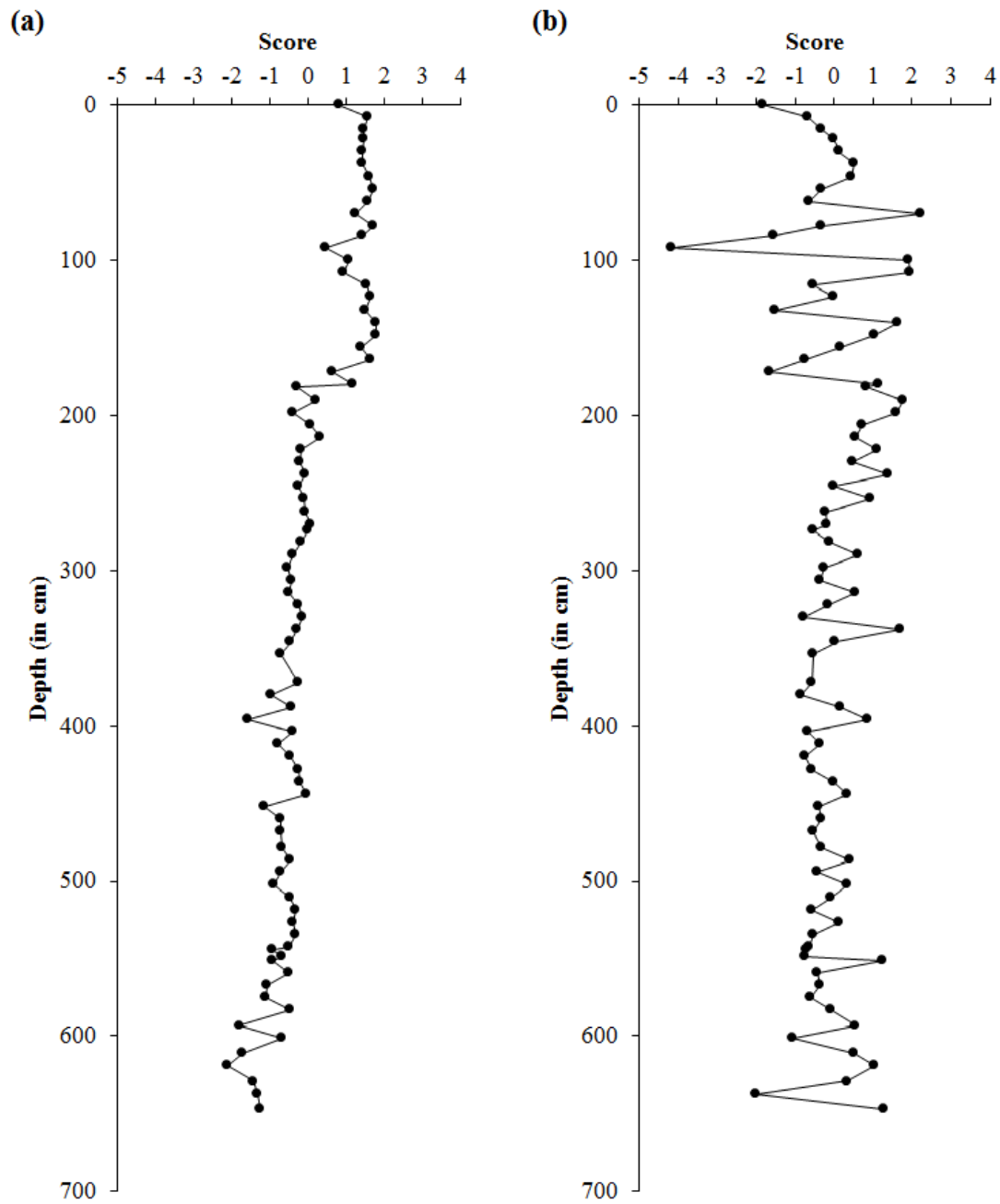


Figure 6



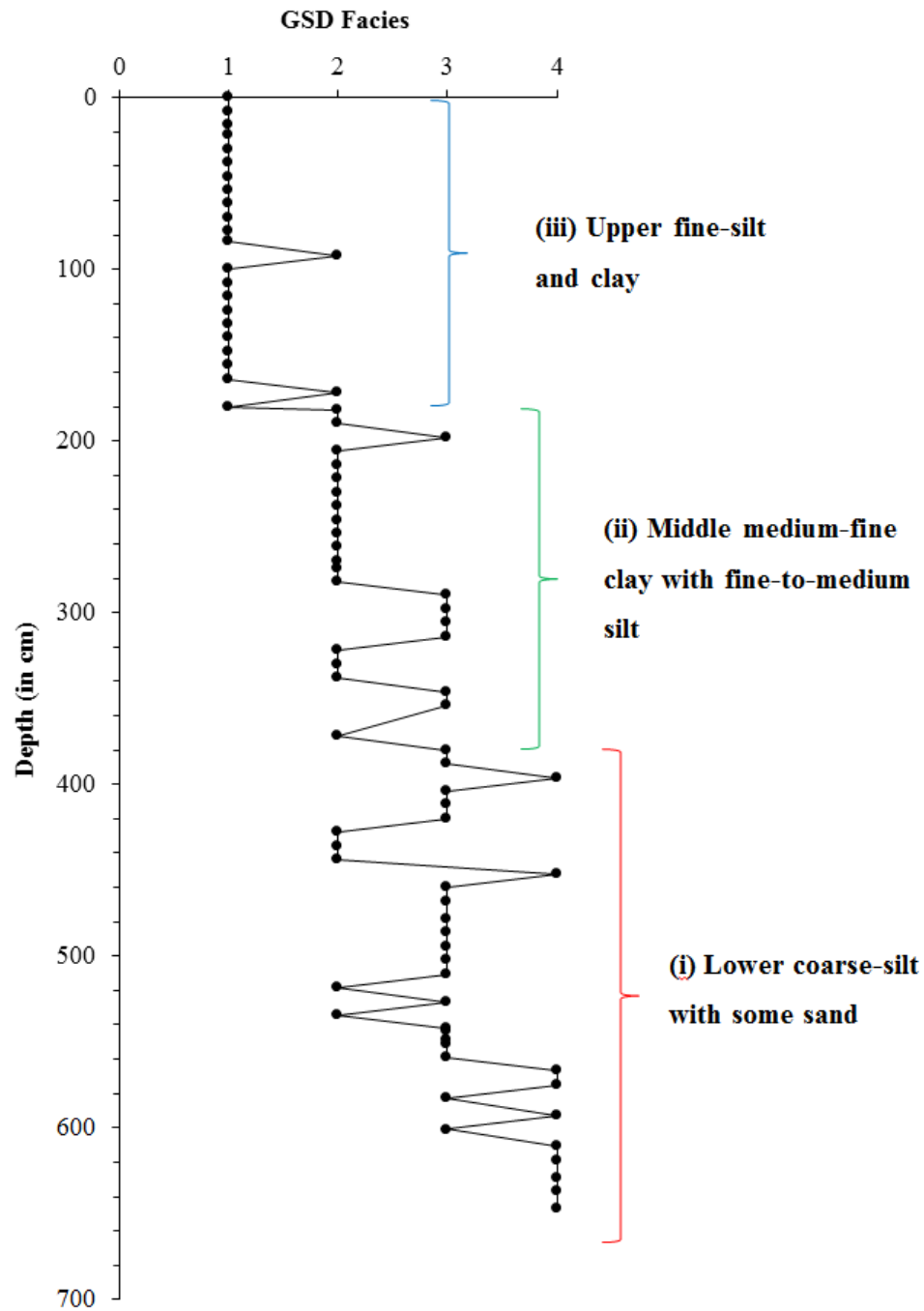


Figure 7

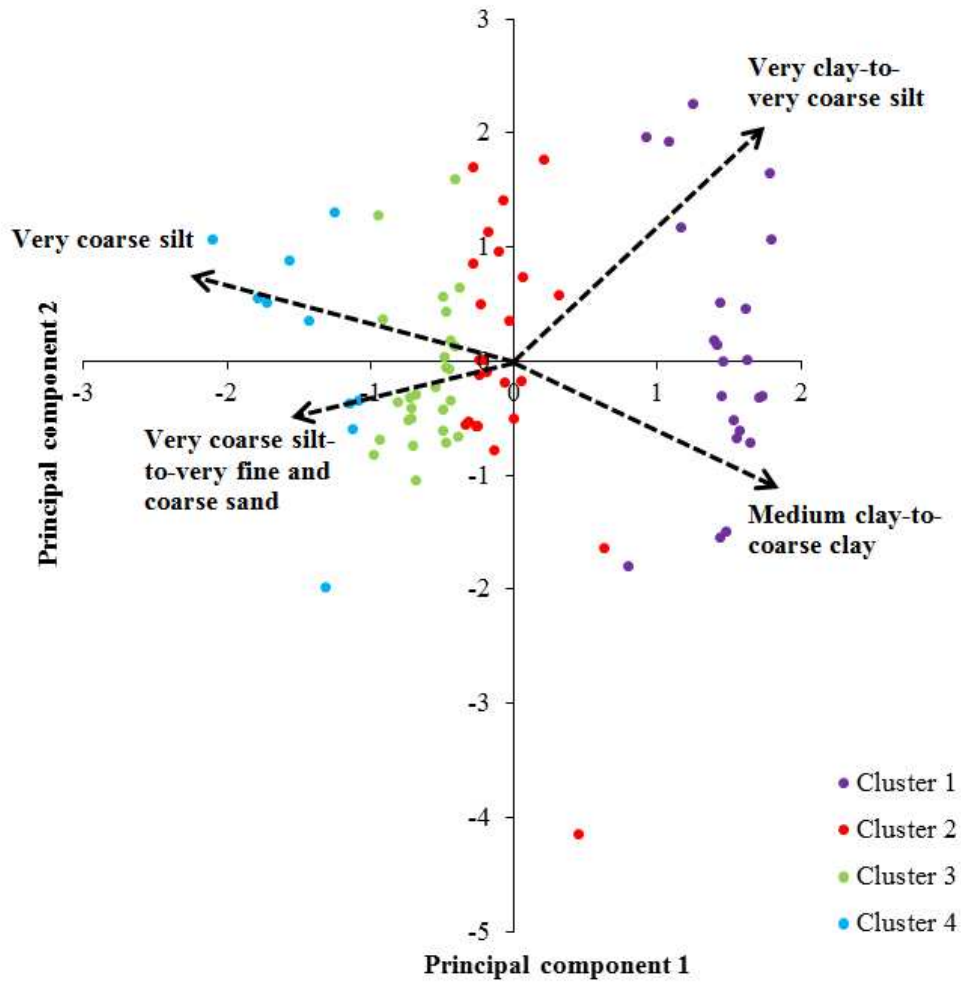


Figure 8

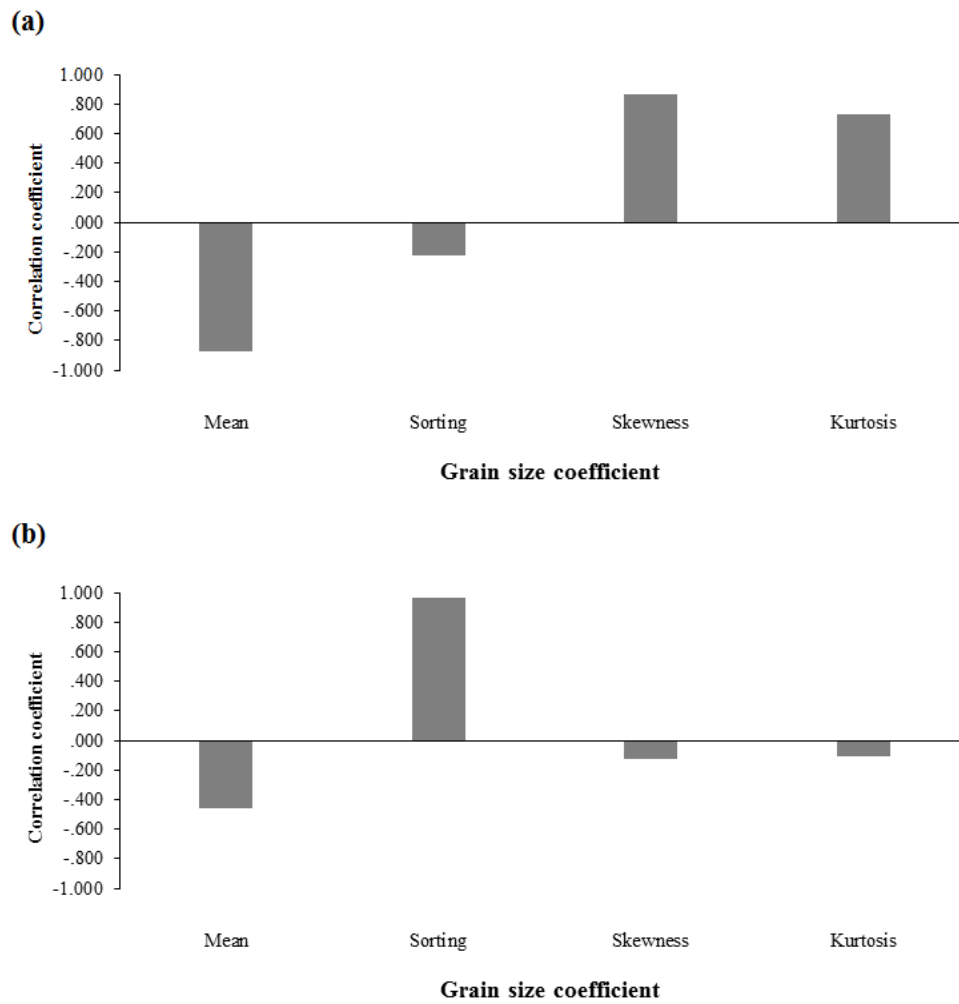


Figure 9

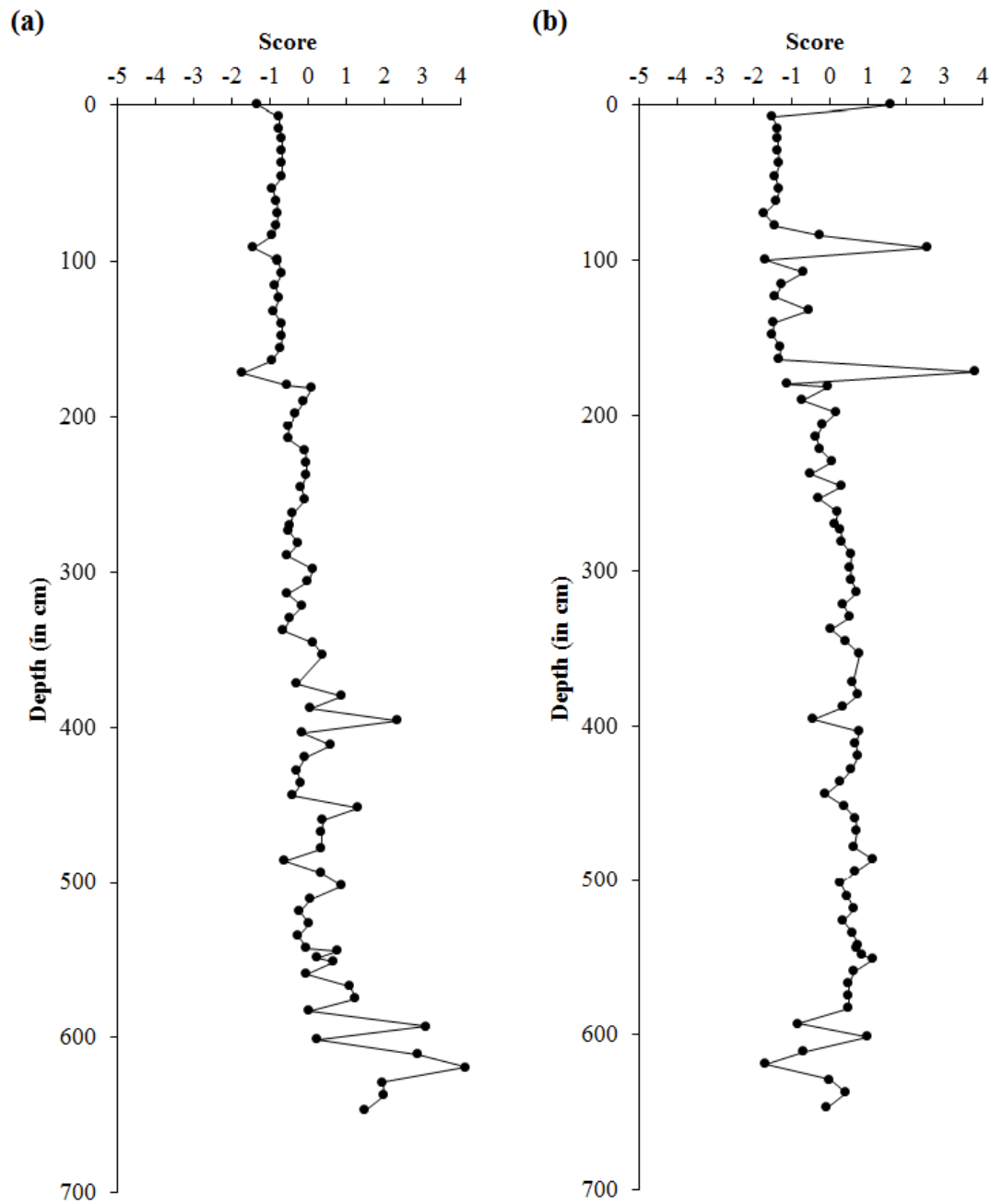


Figure 10

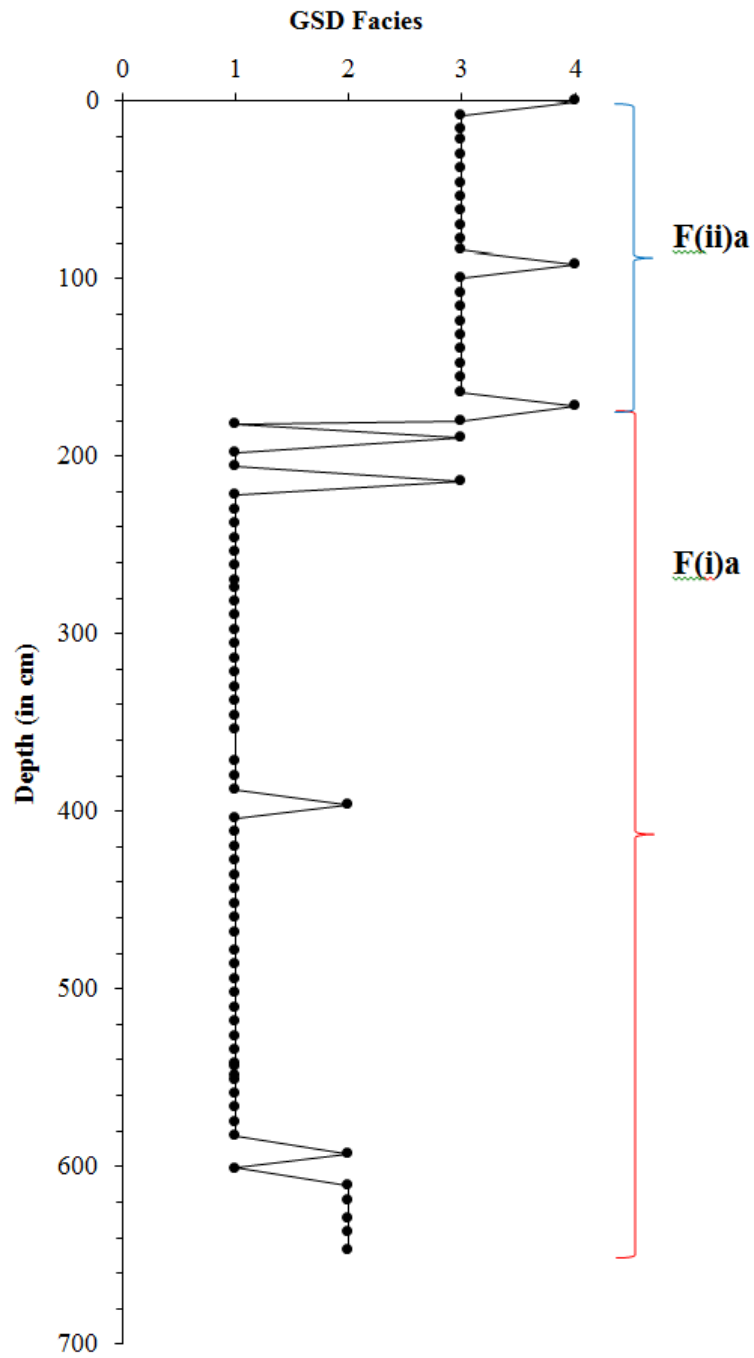


Figure 11

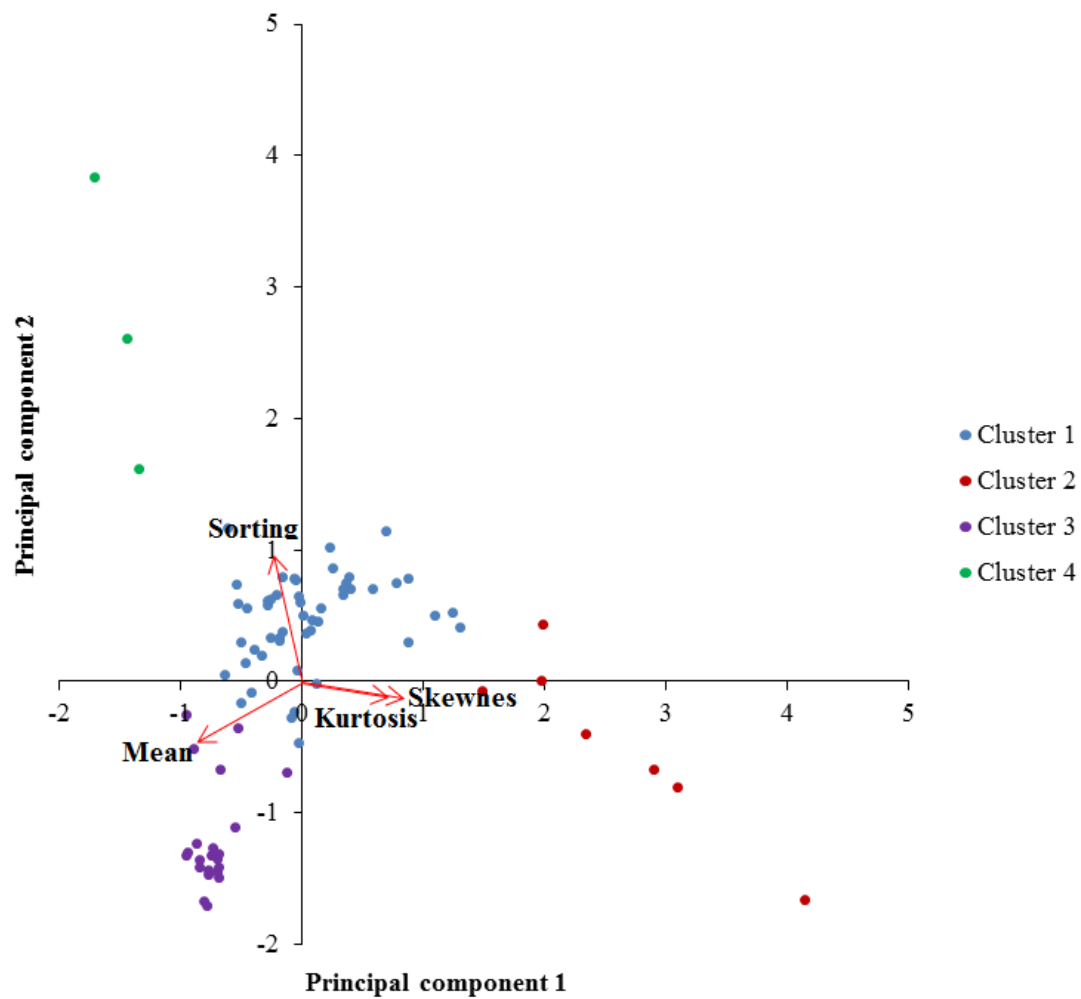


Figure 12

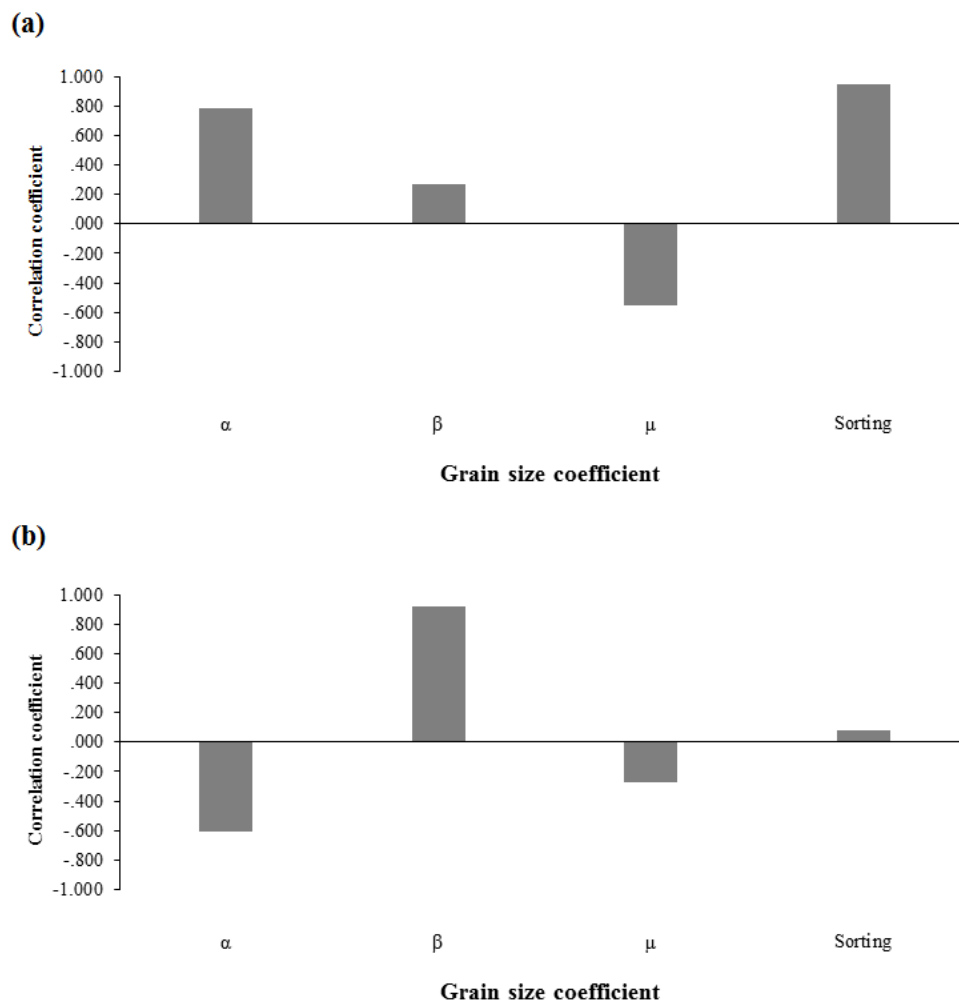


Figure 13

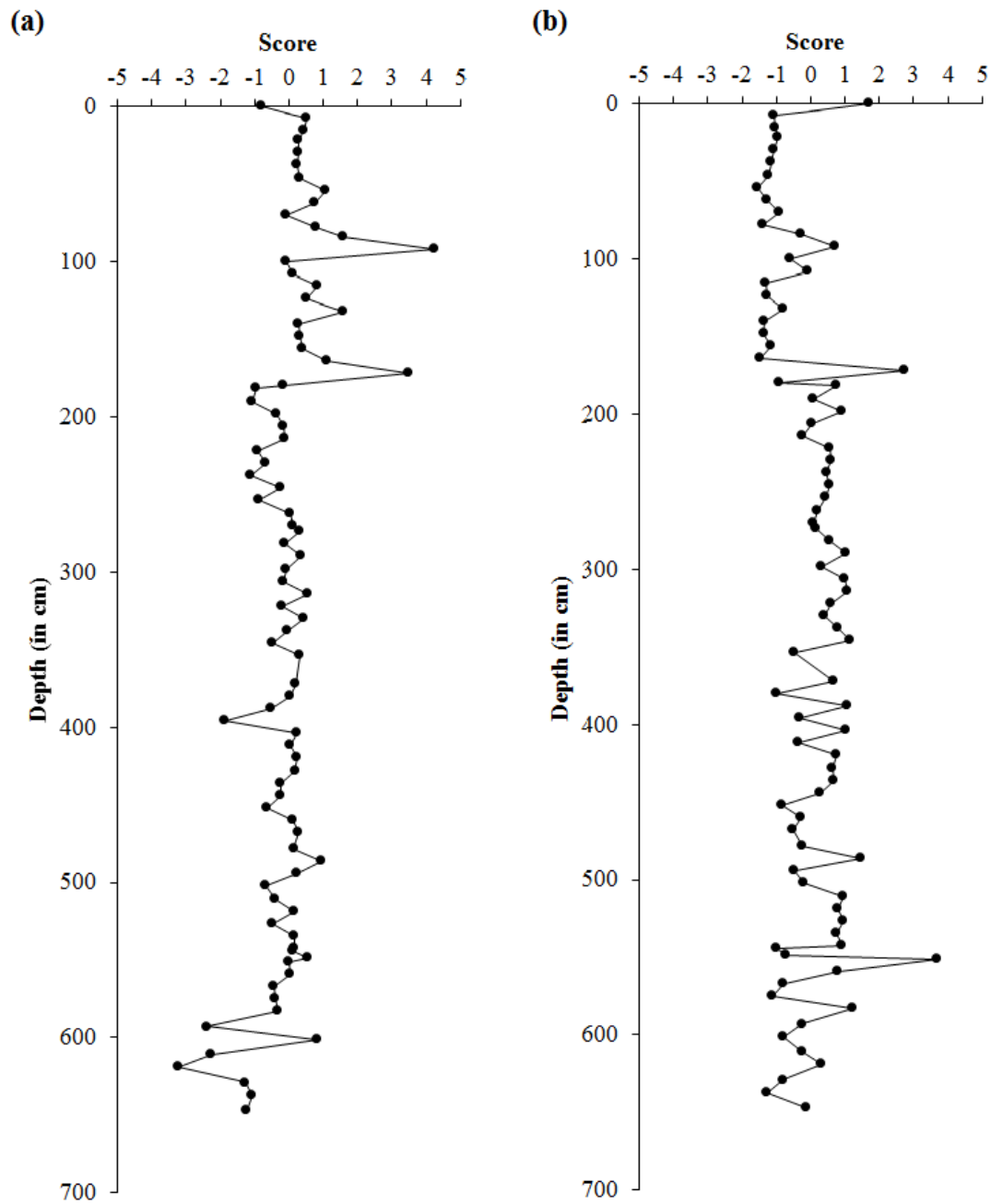


Figure 14



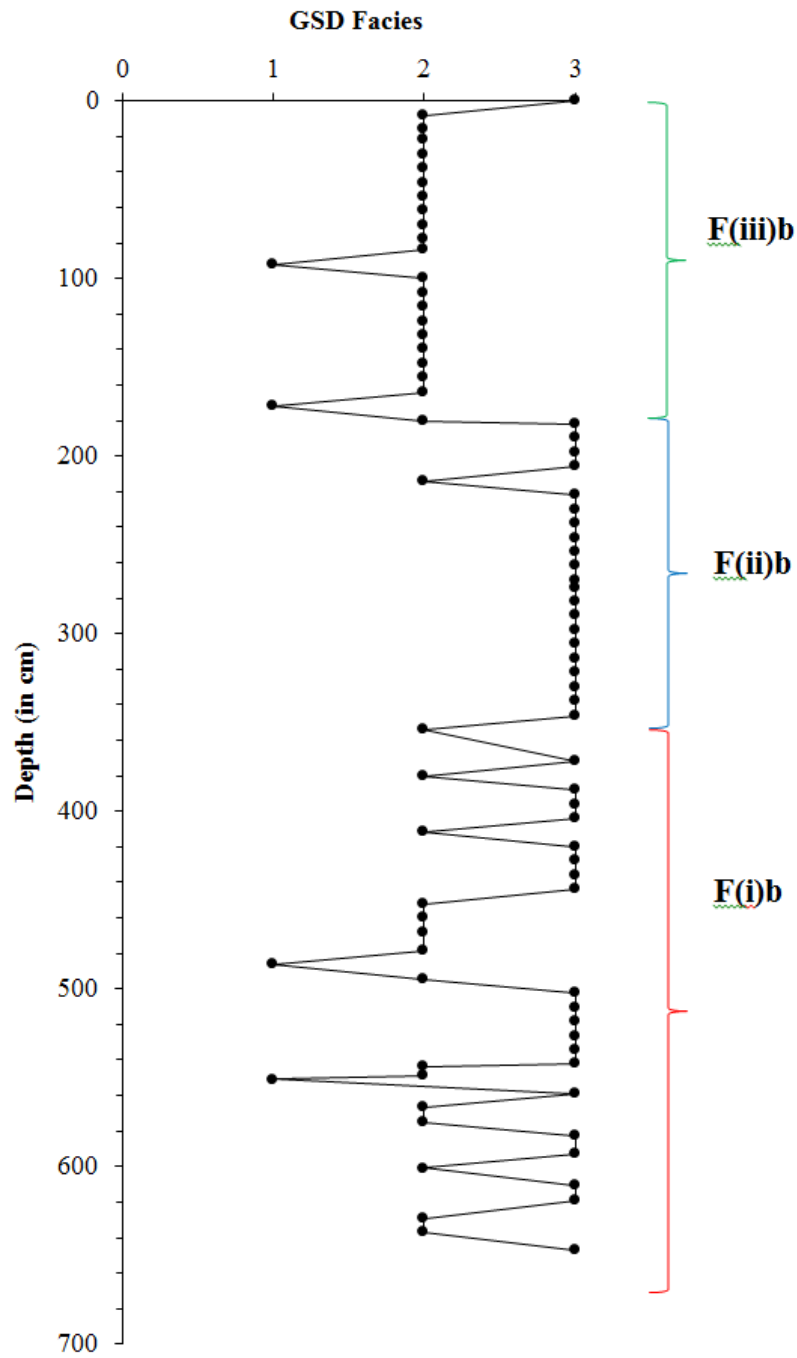


Figure 15

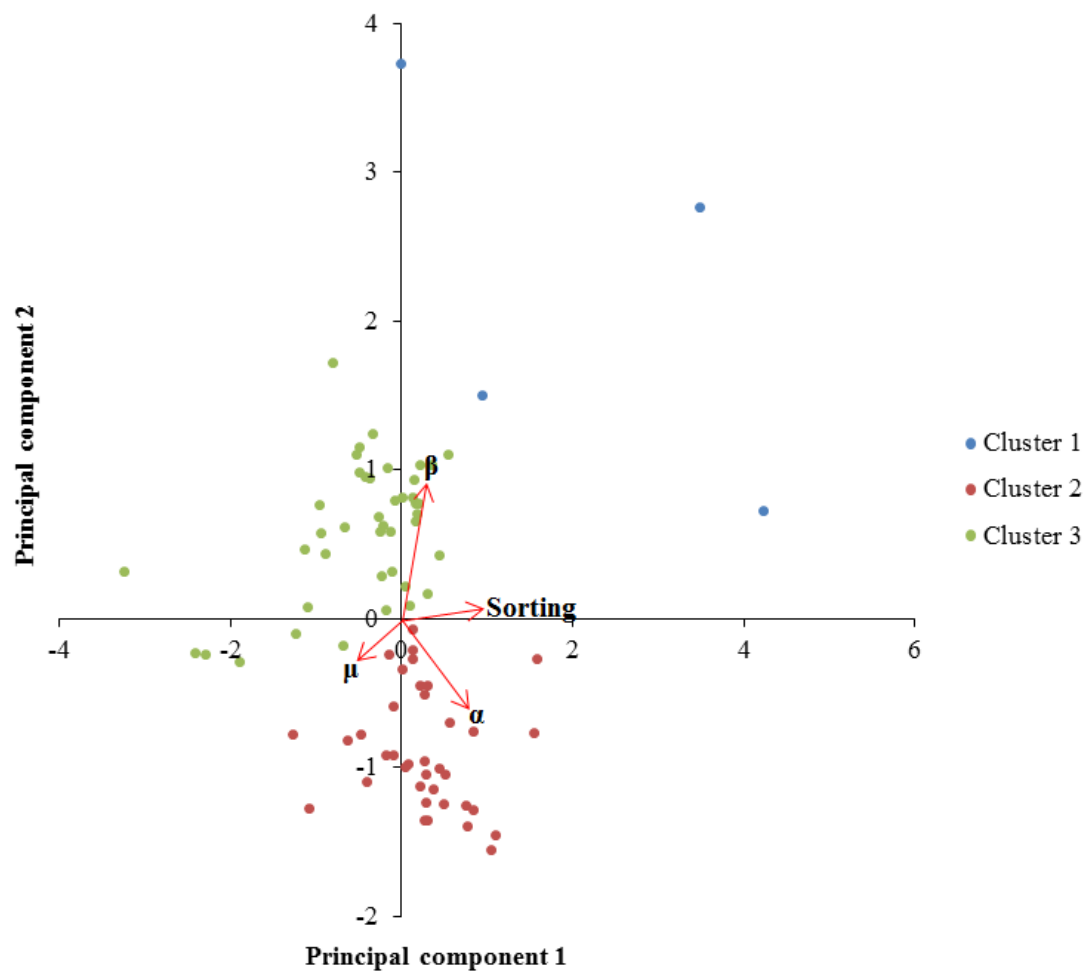


Figure 16

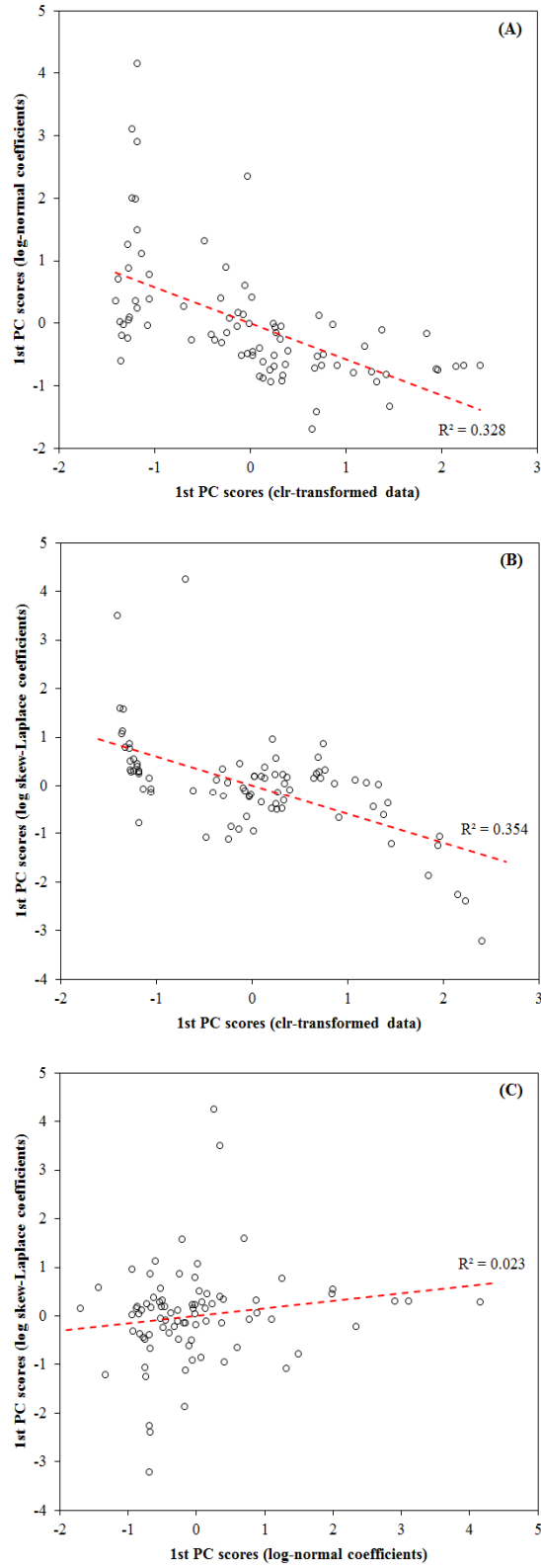


Figure 17

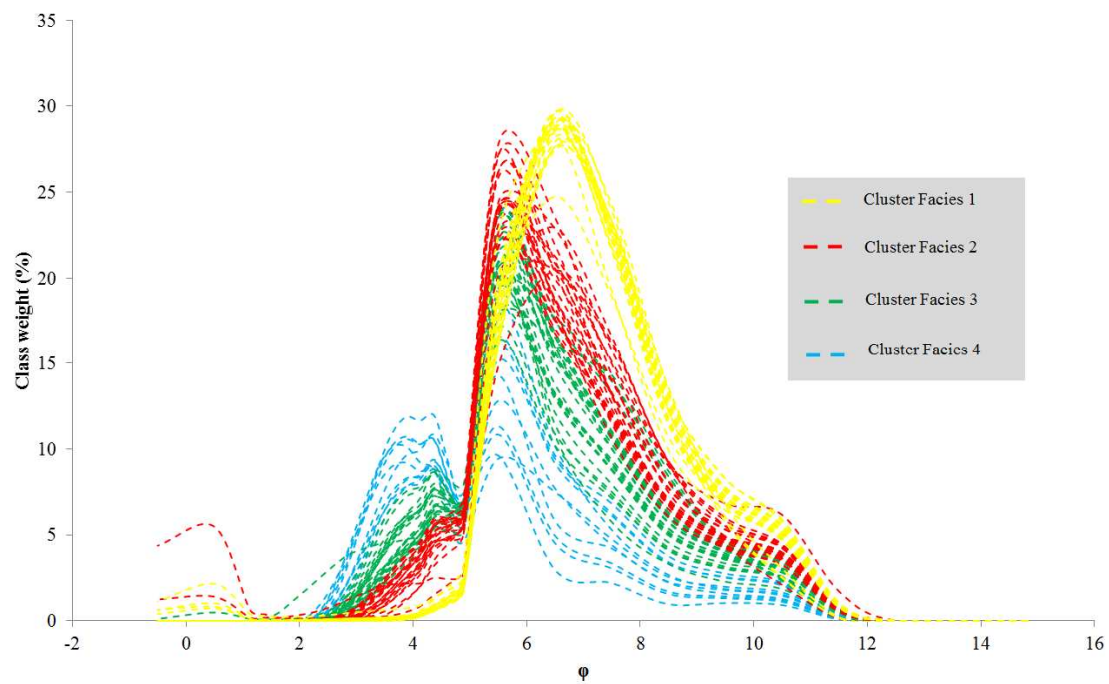
**Figure 18**

Table 1

	12.02φ	10.99φ	9.99φ	8.97	8.00φ	7.00φ	6.00φ	6.01φ	4.76φ	4.61φ	4.24φ	4.00φ
	-	-	-	-	-	-	-	-	-	-	-	-
	10.99φ	9.99φ	8.97φ	8.00φ	7.00φ	6.00φ	6.01φ	4.76φ	4.61φ	4.24φ	4.00φ	-1.00φ
12.02φ												
-	1											
10.99φ												
10.99φ												
- 9.99φ	<b>0.893</b>	1										
9.99φ -												
8.97φ	<b>0.767</b>	<b>0.966</b>	1									
8.97 -												
8.00φ	<b>0.686</b>	<b>0.926</b>	<b>0.991</b>	1								
8.00φ -												
7.00φ	<b>0.687</b>	<b>0.932</b>	<b>0.992</b>	<b>0.997</b>	1							
7.00φ -												
6.00φ	<b>0.708</b>	<b>0.934</b>	<b>0.982</b>	<b>0.987</b>	<b>0.992</b>	1						
6.00φ -												
6.01φ	<b>0.666</b>	<b>0.796</b>	<b>0.838</b>	<b>0.846</b>	<b>0.86</b>	<b>0.899</b>	1					
6.01φ -												
4.76φ	-0.769	-0.89	-0.901	-0.888	-0.88	-0.867	-0.666	1				
4.76φ -												
4.61φ	-0.793	-0.964	-0.971	-0.969	-0.967	-0.948	-0.742	<b>0.971</b>	1			
4.61φ -												
4.24φ	-0.792	-0.969	-0.986	-0.973	-0.976	-0.973	-0.814	<b>0.936</b>	<b>0.993</b>	1		
4.24φ -												
4.00φ	-0.791	-0.974	-0.99	-0.976	-0.98	-0.982	-0.866	<b>0.902</b>	<b>0.978</b>	<b>0.996</b>	1	
4.00φ -												
-1.00φ	-0.776	-0.92	-0.916	-0.898	-0.904	-0.922	-0.904	<b>0.699</b>	<b>0.826</b>	<b>0.873</b>	<b>0.904</b>	1

**Table 2**

	Mean	Sorting	Skewness	Kurtosis
Mean	1			
Sorting	-0.235	1		
Skewness	-0.719	-0.327	1	
Kurtosis	-0.534	-0.231	0.333	1

**Table 3**

	$\alpha$	$\beta$	$\mu$	Sorting
A	1			
B	-0.301	1		
M	-0.18	-0.176	1	
Sorting	0.723	0.403	-0.322	1

**Table 4**

Principal component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	10.818	90.147	90.147	10.818	90.147	90.147
2	0.644	6.367	96.614	0.644	6.367	96.614
3	0.434	3.619	99.133	0.434	3.619	99.133
4	0.076	0.627	99.769	0.076	0.627	99.769
6	0.019	0.16	99.919			
6	0.006	0.046	99.966			
7	0.003	0.026	99.989			
8	0.001	0.007	99.996			
9	0	0.003	99.999			
10	0	0.001	100			
11	0	0	100			
12	0	0	100			

**Table 5**

Principal component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.097	52.423	52.423	2.097	52.423	52.423
2	1.179	29.466	81.889	1.179	29.466	81.889
3	0.683	17.087	98.976			
4	0.041	1.024	100			

**Table 6**

Principal component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	1.898	47.453	47.453	1.898	47.453	47.453
2	1.284	32.104	79.557	1.284	32.104	79.557
3	0.795	19.883	99.439			
4	0.022	0.561	100			

**Table 7**

Dataset	Wald– Wolfowitz Runs Test	Mann-Kendall Rank Test	Cox Stuart test	Bartels Ratio Test (two sided)
First principal component scores: clr-transformed data	$p<0.001$	$p<0.001$	$p<0.001$	$p<0.001$
GSD facies: clr- transformed data (clusters)	$p<0.001$	$p<0.001$	$p<0.001$	$p<0.001$
First principal component scores: log-normal coefficients	$p<0.001$	$p<0.001$	$p<0.001$	$p<0.001$
GSD facies: log- normal data (clusters)	$p<0.001$	$p<0.001$	$p<0.001$	$p<0.001$
First principal component scores: log skew-Laplace coefficients	$p<0.001$	$p<0.001$	$p<0.001$	$p<0.001$
GSD facies: log skew-Laplace data (clusters)	$p<0.001$	$p<0.001$	$p<0.001$	$p<0.001$
Interpretation	alternative hypothesis: non randomness	alternative hypothesis: trend	alternative hypothesis: non randomness	alternative hypothesis: non randomness